Gödel's unpublished papers on foundations of mathematics

W. W. Tait*

Kurt Gödel: Collected Works Volume III [Gödel, 1995] contains a selection from Gödel's Nachlass; it consists of texts of lectures, notes for lectures and manuscripts of papers that for one reason or another Gödel chose not to publish. I will discuss those papers in it that are concerned with the foundations/philosophy of mathematics in relation to his well-known published papers on this topic.[1]

1 Cumulative Type Theory

[*1933o][2] is the handwritten text of a lecture Gödel delivered two years after the publication of his proof of the incompleteness theorems. The problem of giving a foundation for mathematics (i.e. for "the totality of methods actually used by mathematicians"), he said, falls into two parts. The first consists in reducing these proofs to a minimum number of axioms and primitive rules of inference; the second in justifying "in some sense" these axioms and rules. The first part he takes to have been solved "in a perfectly satisfactory way"

*I thank Charles Parsons for an extensive list of comments on the penultimate version of this paper. Almost all of them have led to what I hope are improvements in the text.

[1]This paper is in response to an invitation by the editor of this journal to write a critical review of Volume III; but, with his indulgence and the author's self-indulgence, something else entirely has been produced.

[2]I will adopt the device of the editors of the *Collected Works* of distinguishing the unpublished papers from those published by means of a preceding asterisk, as in " [Gödel, *1933o]". Page references to the papers in volume III will always be to the volume rather than to Gödel's manuscripts. Page references to his published papers, on the other hand, will always be to the original paper (which are indicated in the margins of the first two volumes of *Collected Works* [Gödel, 1986; Gödel, 1990]). Page references alone will always refer to Volume III.

by the formalization of mathematics. Let me summarize what he says about this:

The first attempt at such a formalization, viz. Frege's, in which the axioms and rules of inference "were formulated in a way that seemed to be suggested at first sight", led to obvious contradiction.[3] Only one solution had been found that satisfies the two requirements of avoiding the set theoretic paradoxes and of being adequate to mathematics: the theory of types. He refers here to Russell's simple theory of types, but with three restrictions removed. The first is that no class may be formed containing classes of different types. The second is that the proposition $a \in b$ is to be meaningless when a and b are not of the appropriate types, as opposed to taking it to be meaningful but false. The third restriction to be dropped is the restriction to finite types. When these three restrictions are removed, the theory of types becomes the theory of aggregates as presented by Zermelo, Fraenkel and von Neuman, as becomes clear from [von Neumann, 1929]. Gödel is presumably referring here to von Neumann's proof that the well-founded sets form an inner model of set theory without the Axiom of Foundation. He is saying that, without the above restrictions, the types become the domains $T(\alpha)$, where $T(0) = U$ is the class of urelements, $T(\alpha+1)$ is the domain of subsets of $T(\alpha)$, and for γ a limit, $T(\gamma)$ is the domain of sets of rank γ.[4] The Axiom of Foundation implies that every set has a rank and so a type. Clearly it doesn't matter whether we speak of the hierarchy of $T(\alpha)$'s or the simpler cumulative hierarchy of $R(\alpha)$'s, where $R(\alpha)$ is the set of objects (sets or urelements) of rank $< \alpha$.

But a difference between these hierarchies and Russell's is that the totality of sets obtained in either of the former hierarchies satisfies the Axiom of Extensionality, whereas this is false for the totality of classes obtained in Russell's hierarchy. If $B(n)$ denotes the totality of objects of Russell type n, then $B(0) = U$ and $B(n+1)$ consists of all truth-valued functions over $B(n)$; and so classes of distinct Russell types, having different domains of definition, are distinct. E.g. there is a null class of each type. For $b \in B(0)$, set $b^* = b$

[3]However, in his review [1885] of [Frege, 1884], Cantor had already pointed out that not every extension of a concept is a set. It followed from his theory of transfinite numbers in [Cantor, 1883], as he noted there, that the extension of the concept 'x is a transfinite number' is not a set.

[4]Strangely, Gödel refers to $T(\gamma)$, for limit ordinal γ, as the 'summing up' of the earlier types. In [Gödel, *1939b] he modifies the hierarchy by taking $T(\gamma)$ to be the union of the $T(\alpha)$'s for $\alpha < \gamma$.

and for $b \in B(n+1)$, let b^* be the set $\{x^* \mid x \in B(n) \wedge b(x) = TRUE\}$. Perhaps the way to understand Gödel's assertion that lifting the first two restrictions yields the hierarchy of $T(\alpha)$'s is this: lifting the second restriction gives meaning to propositions involving quantification over all objects (of all types) and in particular to the Axiom of Extensionality $b = c \longleftrightarrow \forall x(x \in b \longleftrightarrow x \in c)$ in Russell's system, even for classes b and c of distinct Russell types. Given $b \in B(n+1)$ and assuming that, on the basis of dropping the two restrictions in question, a^* is defined and $a = a^*$ for all $a \in B(n)$, then since the first restriction is dropped, we can form b^* and on the basis of Extensionality, $b = b^*$. It would follow that $b = b^*$ for all $b \in B(n)$, breaking down the Russell type structure and leaving only the types $T(n)$.

The lifting of the third restriction admits transfinite ranks. Here Gödel notes that Hilbert, for example, had pointed out the possibility of passing to infinite types. No reference is given; but in [Hilbert, 1926] hierarchies of types of $\mathbb{N}$-valued and $\mathbb{N}_1$-valued (as opposed to 2-valued) functions are introduced and extended through the second number class $\mathbb{N}_1$ (i.e. the set of all countable ordinals). There is some internal evidence that Gödel had this work in mind in that, like Hilbert, he takes limit types to be, not the union of the preceding types, but to consist of the functions defined on this union. Since he cites Hilbert only as an example of someone who introduced transfinite types, the question, to which I don't know the answer, arises as to whether there was an earlier introduction of transfinite types. Gödel notes a possible objection to this, which, he suggests, may have been one of the reasons for Russell's restriction to finite types: viz., "in order to state the axioms for a formal system, including all the types up to a given ordinal α, the notion of this ordinal has to be presupposed as known, because it will appear explicitly in the axioms.[5] On the other hand, a satisfactory definition of transfinite ordinals can be obtained only in terms of the very system whose axioms are to be set up." (p. 47) But he believes that this objection can be countered in the following way: "The first two or three types already

[5] In his introductory note to [*1939b; *1940a], Solovay (p. 118) notes that Russell had explicitly rejected transfinite types as impossible. The reason for this was presumably that he wanted the types to be disjoint. But if, as did Hilbert and Gödel, he took the limit type to be, not the union of the earlier types, but the type of functions defined on that union, disjointness would be preserved. A more natural construction, which however would not have suited Russell's purposes, would be to take a limit type $B(\lambda)$ to be the direct limit of the earlier types along the directed family $\{f_{\alpha\beta} \mid \alpha \leq \beta < \lambda\}$, where $f_{\alpha\beta}a$ is defined to be the unique b of type $B(\beta)$ such that $a^* = b^*$.

suffice to define very large ordinals." So we may set up the formal system which formalizes these types, i.e. second or third order predicate logic, "for which purpose no ordinal whatsoever is needed" and in which a transfinite ordinal α_0 can be defined. Then construct the system S_{α_0} which formalizes the hierarchy of $R(\zeta)$ for $\zeta < \alpha_0$. From this system, we can define an even greater ordinal, say β, and construct S_β, "and so on". What does he mean by saying that the first two or three types suffice to define very large ordinals? It is assumed that we begin with an axiomatic theory of a class of urelements, which is provably infinite. We can take it to be the second-order theory of the natural numbers.[6] So at type 1 (i.e. order 2) there is an infinite set $U = \mathbb{N}$ with a definable well-ordering of type ω. As we note below, Gödel is including the second-order Axiom of Choice (i.e. the axiom that the domain of individuals can be well-ordered) among the axioms of type theory. So, even if a set U of urelements were not already given with a definable well-ordering, we could prove the existence of a well-ordering of them of order type $Card(U)$[7] by going up a few types. Presumably, when he speaks of 'defining' an ordinal, he has in mind the proof that such a well-ordering exists. But it is still somewhat puzzling that he speaks of defining 'very large' ordinals in this way. Of course, by going up to still higher types we can in this sense define e.g. the ordinals $\ _n$ for any n, again using Choice (but now in the form that these higher types can be well-ordered).[8] But we will see that, on the one hand, the least upper bound of the hierarchy of ordinals that are obtained starting with any infinite α_0 by the procedure he describes for constructing ordinals [see 1)−2) below] yields (all the ordinals less than) the least strongly inaccessible cardinal greater than α_0 and, on the other hand, his assertion is simply that this procedure will suffice to obtain an inaccessible cardinal. Hence, simply starting with $\alpha_0 = \omega$ would suffice. But perhaps the reference to very large ordinals was only parenthetical, intending to indicate that $\ _n$ for each n can be obtained in this way, without the implication that this is necessary for his construction.

Gödel does not describe the systems S_α, other than to say that they formalize the corresponding hierarchies $\langle R(\zeta) \mid \zeta \leq \alpha \rangle$ (p. 47). But it is possible to fill in details that make sense of the text. Since the hierarchy up to rank α will contain the von Neumann ordinals less than α, we needn't

[6] The axioms are $0 \neq x'$, $x' = y' \longrightarrow x = y$ and second-order mathematical induction.

[7] i.e. a well-ordering of U such that no proper initial segment is equipollent with U.

[8] $\ _0 = \aleph_0$, $\ _{\alpha+1} = 2^{\ _\alpha}$ and for γ a limit ordinal, $\ _\gamma = Lim_{\alpha<\gamma}\ _\alpha$.

treat the ordinals as a distinct sort of object in the formalization, and so may take the universe to consist simply of the urelements and sets. So the language of S_α will contain the unary constant $\mathbf{U}$ for the class of urelements, the relation constant $\in$ and, in order to speak of the hierarchy of $R(\alpha)$'s, a unary function constant denoting the rank function *ran*: $ran(x)$ is the rank of x. (Since we do not want to make any assumptions about α, we cannot assume that the Axiom of Replacement is in general valid in S_α, and so the symbol for *ran* need not be eliminable as a primitive constant in this theory.) There is a second-order theory S_0 whose models (and I will always mean *standard* models, i.e. in which the second-order variables range over all sets of individuals) for a given set of urelements U are, to within isomorphism leaving the elements of U fixed, the structures $M_\alpha = \langle R(\alpha), U, \in, ran\rangle$ for $\alpha > 0$.[9]

It is not unreasonable to suppose that Gödel had in mind for S_α an extension of S_0 such that α is the least ordinal for which M_α is a model of it. That he did have in mind a second-order theory is I think clear, e.g. from the fact that the definition of an ordinal β in S_α of cardinality $> \alpha$ requires the second-order Axiom of Choice. (See the next paragraph.) The S_α are to form an open-ended hierarchy of systems of increasing strength, obtained from S_0 by adding stronger and stronger axioms of infinity, a conception of set theory which is also found in [Zermelo, 1930]. Thus, S_ω would be S_0 together with the axiom that every ordinal has a successor, which is equivalent in S_0 to the Axiom of Unordered Pairs.

When he states that "in terms of the system S_{α_0}" we can define a higher ordinal β and then pass to S_β and "so on", the most reasonable interpretation seems to be that, just as we got $\alpha_0 = \omega = Card(U)$ from the initial class U of urelements, so we proceed from S_α for any α to the next ordinal β by taking $\beta = Card(R(\alpha))$. α can be characterized as the least ζ such that S^ζ_α and so S_β can be obtained by adding to S_0 the axiom that there is a bijection from the ordinals onto the collection of sets of rank $< \alpha$.[10] Notice that, at least on our reading, Gödel is assuming that the Axiom of Choice as a second-order principle (i.e. that $R(\alpha)$ can be well-ordered) is in S_α, both in connection with 'defining very large ordinals' by going up a few types and in connection with defining $Card(R(\alpha))$ in S_α. But in any

[9] S_0 is described in [**?**, p. 274] for the case in which there are no urelements.

[10] S^ζ is obtained by restricting the first-order quantifiers to $R(\zeta)$ and the second-order quantifiers to $R(\zeta + 1)$.

case, Gödel seems to be explicitly including Choice later in the paper when he is speaking of justification of the axioms and lists Choice as one of the weak spots. Moreover, in his past discussions of the simple theory of types, second-order Choice had been included among the axioms, and certainly it is in the set theory of von Neumann, to which he refers, and in the set theory of Zermelo [1930] (to which he does not).

Our reading so far finds some confirmation in Gödel's statement about the place of axiomatic set theory in the hierarchy of systems. He writes

> The place which the system of axioms for the theory of aggregates occupies in this hierarchy can be characterized by a certain closure property as follows: There are two different ways of generating types, the first consisting of going over from a given type to the next one and the second in summing up a transfinite sequence of given types .... Now the statement made by the axioms of the theory of aggregates is essentially this, that these two processes do not lead you out of the system if the second process is applies only to such sequences of types as can be defined within the system itself. [That is to say: If M is a set of ordinals definable in the system and if to each ordinal of M you assign a type contained in the system, then the type obtained by summing up those types is also in the system.] (p. 47)[11]

So we want to consider the least ordinal $\delta > \alpha_0$ closed under these two operations. When he speaks of "going over from a given type to the next one", he can't mean simply ordinal succession: that would have nothing to do with the preceding discussion. Rather, he must mean the passage from α to the β defined in S_α, i.e., to $Card(R(\alpha))$. Since $R(1) = \mathbb{N}$, we have $Card(R(1+\alpha)) = \ {}_\alpha$ for all α. Hence, since $\alpha_0 = \omega < \delta$, we can instead consider the passage from α to $\ {}_\alpha$. Presumably Gödel means by a set M of ordinals definable in the system that it is in $R(\alpha)$ for some α in the system. So, without loss, we could take M to be itself an ordinal and the principle in question is that, if Φ is an operation assigning to each ordinal $< \beta$ an ordinal in the system, where β itself is in the system, then $\bigcup \Phi[\beta]$ is in the system, where $\Phi[\beta] = \{\Phi(\alpha) \mid \alpha < \beta\}$. So his two principles for generating δ from α_0 are

---

[11] Unless otherwise indicated by the appearance of "(W.T.)" after the quote, square brackets and the included text in displayed quotations are Goedel's. In non-displayed quotations, square brackets, as usual, indicate my modification of the quoted text.

1) $\alpha < \delta \longrightarrow \quad {}_{\alpha} < \delta$ (*Jump*)

2) If $\beta < \delta$ and $\Phi : \beta \longrightarrow \delta$, then $\bigcup \Phi[\beta] < \delta$ (*Regularity*)

One thing that is clear from the above passage is that there are, as Feferman suggests in his introductory note (p. 37), two hierarchies involved: one of types or ranks, i.e. ordinals, and the other of theories. For, although we have seen how to get from S_α to S_β for the 'next' type β, obviously no construction of S_α is generally available for the types α introduced by 2). What Gödel has in mind is, first, an autonomous process for generating a hierarchy of bigger and bigger ordinals (or types or ranks) from α_0 by means of 1) and 2) and then, secondly, a sequence $S_{\alpha_0}, S_{\alpha_1}, \ldots$ of increasingly stronger axiomatic theories corresponding to a strictly increasing ω-sequence $\alpha_0, \alpha_1, \ldots$ of ordinals.

It follows from 1)-2) and $\omega < \delta$ that δ is regular and $\delta = \quad {}_{\delta}$, i.e. that δ is strongly inaccessible. We will denote the least $\delta > \alpha$ which satisfies 1)-2) by $F(\alpha)$. So for $\delta = F(\omega)$, the system S_δ can indeed be taken to be "the system of axioms for the theory of aggregates", M_δ being its least (standard) model. We have here an account of 'cumulative' type theory or the 'iterative concept of set', which Gödel introduced in print in [1947; 1964]. It is worth emphasizing that, for Gödel, both in this earlier unpublished paper, in the unpublished [*1951] and in the later published work, as opposed to [Shoenfield, 1967] or [Boolos, 1971], for example, the hierarchy of types is *autonomous*: it does not presuppose an externally given system of ordinals or 'stages'.

However, Gödel is not concerned only with hierarchies generated by 1) and 2), yielding $F(\alpha)$ from α. For example, replacing 1) by the new jump

$$\beta < \delta \longrightarrow F(\beta) < \delta$$

we obtain for any α an inaccessible cardinal $\delta = G(\alpha) > \alpha$ which is the limit of inaccessible cardinals. And so on:

> There is no end to this process [and the totality of all systems thus obtained seems to form a totality of similar character to the set of ordinals of the second number class].[12] (p. 47)

[12]The reference to the second number class in the bracketed clause is a puzzle. It seems to refer to the fact that the second number class is closed under least upper bounds of countable sets of ordinals of the second number class. But clearly an increasing sequence

It should be noted, as it effectively already was in [Zermelo, 1930], that Gödel's concern about constructing the infinite ordinal α_0 to begin the process was unnecessary: closing 0 under the jump operation $\alpha \mapsto \alpha + 1$ yields ω. So, in particular, the assumption of an infinite class of urelements is not needed to begin the hierarchy. In [*1951] Gödel again describes the cumulative hierarchy, indicating that the Jump can be quite arbitrary: it can be

$$\alpha < \delta \longrightarrow H(\alpha) < \delta$$

for any given ordinal-valued function H of ordinals.[13]

It is easy to see that each of these closure conditions, both under the various jumps and under the Regularity operation, has the form of the *reflection principle*

$$\forall X[\phi(X) \longrightarrow \exists\beta\phi^\beta(X)]$$

where $\phi(X)$ is a first-order formula of set theory containing only the second-order variable X (i.e. ranging over classes) free. In fact, M_δ satisfies the reflection principle for all first-order $\phi(X)$ just in case δ is inaccessible. Applied to Π^1_1 formulas $\phi(X)$,[14] the reflection principle was first used in [Lévy, 1960] to derive Mahlo cardinals in set theory. Gödel cites this paper in [1964] and, in particular, in footnote 20 of [Gödel, 1964],[15] as establishing the existence of Mahlo cardinals on the iterative conception. In fact, although Gödel didn't mention it, the condition that M_κ satisfy the reflection principle applied to Π^1_1 formulas $\phi(X)$ in general was proved equivalent to κ being a weakly compact cardinal in [Hanf and Scott, 1961].

$\langle S_{\alpha_n} \mid n < \omega\rangle$ is going to be bounded by a system S_α only if M_α satisfies some axiom which implies all of the S_{α_n}. In particular, there can be only a countable number of S_α's. On the other hand, he could be referring the other hierarchy, of the ordinals themselves, noting that this hierarchy is closed under least upper bounds of arbitrary sets of ordinals. But there is a difference between this hierarchy and the second number class, in that, whereas, among classes of ordinals, the notion of a *countable set* of ordinals is mathematically characterizable, the notion of a *set* of ordinals is not: this is precisely why the sequence of ordinals and the sequence of theories S_α are open-ended.

[13]He does not explicitly mention the Regularity operation 2) here; he speaks only of iterating the jump 'into the transfinite'. But, even when the jump is just $\alpha \mapsto \alpha + 1$, taking arbitrary iterations, without the restriction at limit stages given in 2), yields the class of 'all' ordinals, presupposing that this class is externally given.

[14]Π^1_1 formulas are those of the form $\forall Y\psi(X,Y)$ where $\psi(X,Y)$ is first-order and Y ranges over classes.

[15]This is a revised footnote, carrying the date: September 1966.

But note that there is a conceptual difference between the above reflection principle when $\phi(X)$ is first-order and when it is higher-order. When it is first-order, the principle simply asserts, for any class A such that $\phi(A)$, the existence of an ordinal closed under the Skolem functions for $\phi(A)$, which is closure under a jump in the above sense. When $\phi(X)$ is Π^1_1, say $\phi(X) = \forall Y \psi(X, Y)$ where $\psi(X, Y)$ is first-order, the principle asserts that, for any class A such that $\phi(A)$, there is an ordinal β such that $R(\beta)$ is closed under the Skolem functions $\Phi(B)$ for $\psi(A, B)$ for *all* classes $B \subseteq R(\beta)$. Here the reflection is not on closure under a jump, but on a general logical condition that the universe satisfies, and the principle asserts that, then, there is some ordinal β such that $R(\beta)$ also satisfies it. One might imagine that, even if the existence of ordinals closed under given functions seems a compelling axiom, these stronger forms of the reflection principle might become qualitatively less compelling as $\phi(X)$ becomes logically more complex.

In the footnote cited above, Gödel mentions certain stronger axioms of infinity (than the existence of Mahlo or weakly compact cardinals), but asserts "That these axioms are implied by the general concept of set in the same sense as Mahlo's has not been made clear yet ...". The example that he mentions asserts the existence of a measurable cardinal, which implies that not all sets are constructible [Scott, 1961]. It remains unclear even now whether the 'general concept of set' in question implies the existence of non-constructible sets. In his introductory note to [*1939b; *1940a], Solovay calls into question Gödel's statement about his proof of the consistency of $V = L$ that "the consistency therefore holds in the absolute sense, insofar as it makes any sense at all today to speak of absolute consistency" (p. 129), on grounds of the existence of large cardinal axioms which are inconsistent with $V = L$. But Gödel's "therefore" refers to the fact that his proof that $V = L$ is consistent remains valid under the addition of "new *evident* axioms" (my emphasis), which, even in his later papers on the subject, did not include the existence of measurable cardinals. The evidence that would seem to be demanded by him here is that the proposed axiom assert the existence of a cardinal with some property already possessed by the totality of all ordinals. (However, as [Reinhardt, 1974] points out, without a careful analysis of what one is to mean here by such a property, this principle for introducing new axioms leads immediately to contradiction.) Solovay [Gödel, 1990, p. 19] mentions that Gödel later on believed in the existence of measurable cardinals; but the evidence for this belief is of a different—one might say *opposite*—sort: a property possessed by ω should be shared by some bigger cardinal. (It is even harder to circumscribe the notion of property in this context to avoid inconsistency.)

Returning to [*1933o], it is the formalization of one stage or another in the hierarchy of M_α's that constitutes for Gödel the solution to the first part of the problem of foundations. That there is no one such formal system, but rather an open ended hierarchy of them, does not for him detract from it being a "perfectly satisfactory" solution. For one thing, "all the methods and proofs hitherto developed are in [Zermelo-Fraenkel set theory], and, apart from certain theorems of the theory of aggregates, all mathematics hitherto developed is contained even in much weaker systems, which include just a few of the first types" (p. 48). But, also, he thinks that the open-endedness of the theory of types is a "strong argument in [its] favor", because it fits well with the phenomenon of incompleteness which he discovered for formal systems in general and which we mentioned above: applied to a consistent system $S = S_\alpha$, where α is infinite, there is a "proposition in the arithmetic of integers" which is provable " if you add to the system S the next higher type and the axioms concerning it", in other words, it is provable in $S+\exists\zeta S^\zeta$, but not provable in S.[16] If S is an extension of second-order set theory, for example, then α is an inaccessible cardinal and $S + \exists\zeta S^\zeta$ is satisfied first in M_β, where β is the next greatest inaccessible cardinal. Presumably, the open-endedness of the cumulative type theory is a strong argument in its favor because the 'natural' proof of the undecidable proposition, i.e. the proof underlying our conviction that it is *true*, involves the notion of truth in S and so requires the system $S + \exists\zeta S^\zeta$.

Of course, we already obtain $S_\beta = S + \exists\zeta S^\zeta$ in the cumulative hierarchy from $S = S_\alpha$ by reflection on the sentence $\phi(X) = S$. But Gödel points out that his incompleteness theorems are of interest from another point of view regarding this hierarchy: we have a proof of an arithmetic proposition in S_β not obtainable in S_α, and so "the construction of higher and higher types is by no means idle ...'' (p. 48). This result, to which Gödel refers back in his later works (including [*1951] as well as his published works), is

[16]Based on his result in [*193?], Gödel sharpens his statement in [*1951] by noting that the undecidable proposition in question can be taken to be "the solution of certain diophantine problems" (p. 307), by which he means it has the form

$$\forall x_1 \dots x_m \exists y_1 \dots y_n P(x_1, \dots, y_n) = 0$$

where P is a polynomial.

especially pertinent in view of the emphasis in [*1933o] on the 'totality of methods actually used by mathematicians". It points to the possibility of validating the existence of a large transfinite number by showing that it has low-down implications, for example concerning the integers. This subject has been taken up in more recent years, especially by H. Friedman.[17]

However, in [*1951] Gödel remarks that it 'is a mere historical accident, which is of no importance for questions of principle" that "99.9% of present-day mathematics is contained in the first three levels of [the cumulative] hierarchy" (p. 307). The open-ended hierarchy of types and, in particular, large cardinals do not need for their justification low-down applications: they are intrinsic to the concept of transfinite number as this notion was introduced in [Cantor, 1883]. That is presumably what Gödel meant by speaking of "questions of principle". In this respect, also, the view drawn by some writers from the later papers [Gödel, 1947; Gödel, 1964], that Gödel's concern for new axioms of set theory was motivated by the desire to settle arithmetic problems or the continuum problem,[18] is at best misleading: whatever their relation to this problem, the need for ever new axioms is intrinsic to Cantor's theory of transfinite numbers, with the principle that every set of numbers has an upper bound.

As I have already mentioned in [?], Gödel actually introduces a criterion for accepting new axioms which is not necessarily compatible with the idea of the hierarchy arising from the idea that every 'property' of the totality of ordinals must be possessed by some ordinal. On p. 265 of [1964] he writes

> Secondly, however, even disregarding the intrinsic necessity of some new axiom, and even in case it had no intrinsic necessity at all, a probable decision about its truth is possible also in another way, namely, inductively by studying its "success". Success here means fruitfulness in consequences, in particular in "verifiable" consequences, i.e., consequences demonstrable without the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover, and make it possible to condense into one proof many different proofs.

[17]See for example [Friedman, 1998]. There is a fair amount of unpublished work by Friedman on this subject, to be found at his website: www.math.ohio-state.edu/foundations/manuscripts.html.

[18]For example, see [Feferman, 1997]. Feferman locates the motivation for Gödel's "program for new axioms" in his discovery of the incompleteness of formal systems, ignoring the intrinsic necessity for new axioms arising from the notion of transfinite number itself.

It is difficult to reconcile this with the iterative conception of the universe of sets we are discussing here. On the latter conception, the 'intrinsic necessity' of an axiom arises from the fact that it expresses that some property possessed by the totality of ordinals is possessed by some ordinal. To introduce a new axiom as 'true' on this conception because of its 'success', would have no more justification than introducing in the study of Euclidean space points and lines at infinity because of their success. One may obtain an interesting theory in this way and one worthy of study; but it won't be Euclidean geometry. A 'probable decision' about the truth of a proposition from the point of view of the iterative conception can only be a probable decision about its derivability from that conception. Otherwise, how can we know that a probable decision on the basis of success might not lead us to negate what we otherwise take to be an intrinsically necessary truth?

2 Impredicativity

The second part of the problem of foundations according to [Gödel, *1933o] is that of giving a justification for the axioms and rules of inference. Gödel writes concerning this problem that "the situation is extremely unsatisfactory" (p. 49). He cites three kinds of difficulties or "weak spots": i) The use of the non-constructive notion of existence in connection with 'surveyable' objects such as the integers. In particular, we apply the law of excluded middle to existence statements, yielding proofs of the existence of a number, for example, with a given property although the proof yields no algorithm for finding such a number and, indeed, we may have no clue at all as to how to find one, "just as if in some objective realm of ideas this question [of existence or not] were settled quite independently of any human knowledge" (p. 49). ii) The second weak spot, which is "still more serious", is the treatment of the notion of class in set theory; and iii) the third concerns the Axiom of Choice.

Concerning iii), he writes that he will not go into details because "it is of less importance for the development of mathematics." In view of the role that Choice plays in generating the hierarchy of types, this is a surprising remark: without this axiom, how does one 'define' very large ordinals by going up two or three types and how, in S_α, does one in general 'define' ordinals of cardinality greater than α? Perhaps the solution to the problem posed by Gödel's remark—at any rate, I have not been able to think of another one—is that suggested by Feferman in his introductory note: he points out (p. 39) that Gödel had established the relative consistency of the axiom of choice in Zermelo-Fraenkel set theory in 1935 (although he only announced the result in 1938) and suggests that he may have been already thinking along these lines in 1933. In fact, [Wang, 1987, 97] recounts Gödel's recollection in 1976 that he thought about the continuum problem from around 1930 and that the idea of using the ramified hierarchy occurred to him quite early: the problem was to construct a big enough class of ordinals. It was the use of the classical (impredicative) ordinals, presumably in 1935, that yielded his result. But why, given his attitude towards impredicative definition in [*1933o] (see below), did he feel in position in 1933 to wave aside the problem of the Axiom of Choice but not the problem of impredicative definition?

In connection with ii), one problem concerns the non-constructive notion of existence, e.g., the application of the law of excluded middle to propositions asserting the existence of sets. But also he is concerned about the use of impredicative definition of sets (or, what he takes to be the same, properties): in S_α impredicative definition of sets (in $R(\alpha)$) is admitted. He writes

> Again, as in the case of the law of excluded middle, this process of definition presupposes that the totality of all properties exists somehow independently of our knowledge and our definitions, and that our definitions merely serve to pick out certain of these previously existing properties. If we assume this, the method of non-predicative definition is perfectly all right .... But the situation becomes entirely different if we regard the properties as *generated* by our definitions. For it is certainly a vicious circle to generate an object by reference to a totality in which this very object is supposed to be already present. (p. 50)

Thus, Gödel subscribed at this time to the vicious circle principle in the sense that he regarded what he refers to as 'Platonism' as its only alternative and, as we will see, he was critical of Platonism; but he seems to have changed his mind on the question of whether the vicious circle principle implies Platonism by the time he composed [*1938a] and he certainly changed his attitude towards Platonism by the time he wrote [1944].

In [*1938a, p. 94], in discussing what constraints should be imposed on extensions of primitive recursive arithmetic in order that they meet the demands of constructivity, Gödel not only imposes no restriction to predicative definitions, but he introduces as the extension best satisfying those constraints a theory of computable functions of finite type, subsequently developed in [*1941] and [1958], and he explicitly admits in this theory impredicative definitions of functions—e.g. defining a function f of some type A in terms of a function Φ of type $A \longrightarrow \mathbb{N}$ (p. 95). The finite types can be taken to be built up from $\mathbb{N}$ by means of the operation $\longrightarrow$, where $A \longrightarrow B$ is the type of the calculable functions from A to B. We will discuss this theory below in §5; but the thing to be noticed here is that, by 1938, he was willing to accept impredicative definition of functions as constructive. Gödel remarks that this is not circular, because the functions are computable. Thus, on this conception, functions do not enter in as objects, but only through their roles in the computation, ultimately, of numerical terms. But it is certainly so that, even though the value $\Phi(f)$ may be defined by a computation without reference to the whole function Φ, the element $\langle f, \phi(f)\rangle$ of the graph of Φ in general cannot; and so on the conception of functions as graphs, the vicious circle principle is literally violated. We will discuss the question of the computability of these functions in §5 and will see that there is a circle, too, in the argument for their computability.

In [1944], in response to the necessity for impredicative definitions in classical mathematics, he writes

> I would consider this rather as a proof that the vicious circle principle is false than that classical mathematics is false, and this is indeed plausible also on its own account. For, first of all one may, on good grounds, deny that reference to a totality necessarily implies reference to all single elements of it or, in other words, that 'all' means the same as an infinite logical conjunction. (pp. 135-6)

In order to understand the "in other words" here, it should be noted that in ramified type theory, finite or transfinite, as opposed to simple type theory, we can assign ordinal ranks to formulas in such a way that the rank of a formula is always greater than the ranks of its *components*. The components of $\phi \wedge \psi$ and $\phi \vee \psi$ are ϕ and ψ and the components of $\forall\zeta\phi(\zeta)$ and $\exists\zeta\phi(\zeta)$ are the formulas $\phi(\tau)$, where τ is a term of the appropriate sort. We can assume that the logical constants are $\wedge$, $\vee$, $\forall$ and $\exists$ and that negations are applied only to atomic formulas. Thus, $\neg\phi$ in general is obtained by successively moving negations inside the scope of other logical constants, using the De Morgan laws, and dropping double negations. $\neg\phi$ has the same rank as ϕ.

A $\wedge$- or $\forall$-formula, called a $\bigwedge$-*formula*, is thus expressed by the conjunction of its components and a $\vee$- or $\exists$-formula, called a $\bigvee$-*formula*, is expressed by the disjunction of them; and so, by induction on its rank, it follows that every sentence can be represented as an infinitary propositional formula built up from atomic and negated atomic sentences concerning the urelements.

Gödel then goes on in [1944] to essentially repeat what he wrote in the quotation above from [*1933o]: "all" can mean infinite conjunction, even in the presence of impredicative definitions, providing the objects in question are understood to be *sui generis* and not our creatures. Thus, the τ such that $\psi(\tau)$ is a component of $\forall X\phi(X)$ and $\exists X\phi(X)$, where X ranges over sets of numbers, for example, are not definitions, but simply all of the names of sets of numbers, so that the formulas $n \in \tau$ and $n \notin \tau$ are taken to be atomic sentences with well-defined truth-values. In this case, too, every formula will have a rank such that the step from formula to component always lowers rank: the rank of a formula is just the finite maximum number of nested logical constants in it. So the formulas are 'infinitary propositional formulas'; but of course in this case they are not really syntactical objects, since their atomic parts involve the names of arbitrary sets of numbers.

In analogy with Gödel's attitude towards his functions of finite type, one might feel that the vicious circle principle is not the issue. Just as in the case of these functions, sets may be defined by formulas involving quantification over their types: we only need to know that from those definitions the truth-values of sentences involving these quantifications are determined. In the above cases in which "all" means conjunction, this is easily seen. We define the game $\mathcal{T}(\phi)$ played by $\bigvee$ and $\bigwedge$. Each stage of the game consists of exactly one sentence on the board, with ϕ on the board at the first stage. Let ψ be on the board at a given stage. If ψ is an atomic or negated atomic sentence, then the game is over and is won by $\bigvee$ if the sentence is true and by $\bigwedge$ otherwise. If ψ is a $\bigwedge$-sentence, then $\bigwedge$ chooses one of its components and puts it on the board at the next stage; and if it is a $\bigvee$-sentence, then $\bigvee$ chooses one of its components and puts it on the board at the next stage. The truth of ϕ consists in the existence of a winning strategy for $\bigvee$ in $\mathcal{T}(\phi)$. Since successive moves in $\mathcal{T}(\phi)$ lower the rank of the sentence on the board, all games are all finite: there are no draws. By induction on the height of the well-founded tree of all games, i.e. the rank of ϕ, one easily shows that $\bigvee$ has a winning strategy in precisely one of $\mathcal{T}(\phi)$ and $\mathcal{T}(\neg\phi)$; and so, in particular, $\phi \vee \neg\phi$ is true.[19]

But if we try to apply this conception to impredicative mathematics, say to number theory of some finite or infinite order > 1, where the components of $\forall X\phi(X)$ and $\exists X\phi(X)$ are $\phi(\tau)$, where τ is a closed term $\{x \mid \psi(x)\}$, we notice, first, that "all" cannot mean conjunction: for example, let $\phi(X)$ be $0 \in X$, where X is a second-order variable, and let $\tau = \{y \mid \forall X(y \in X)\}$. Then the attempt to represent $\forall X\phi(X)$ as the conjunction of all of its components is circular, since the component $\phi(\tau)$ is just $\forall X\phi(X)$ itself. (I am treating the expression $t \in \{x \mid \phi(x)\}$ as an abbreviation for the corresponding $\phi(t)$.) It then follows that there are in general infinite games of $\mathcal{T}(\phi)$ in this system, as when ϕ is $\forall X(0 \in X)$ and dumb $\bigwedge$ keeps choosing ϕ itself at each stage. There is indeed a circle here; and the effect of such circles or, more generally, of infinite games is that it is no longer possible, by an elementary proof such as we have in the predicative case, to establish the logical law $\phi \vee \neg\phi$. To establish this, we need to invoke 'Platonism', at least to the extent of assuming the existence of a (Henkin) ω-model for the theory in which every set is defined by a formula of the theory. We might think to replace the game $\mathcal{T}(\phi)$ with Gentzen's game $\mathcal{G}(\{\phi\})$, described below in §7, in the definition of truth, since with this definition, the validity of the law of excluded middle becomes trivial. The difficulty is that another logical law, *modus ponens*, requires this same 'Platonistic' assumption.

In connection with the "good grounds" for the alternative of rejecting logical conjunction as the meaning of "all", Gödel mentions as an example only the suggestion of Carnap and Langford that "all" means "analyticity or necessity or demonstrability". He indicates that there are difficulties with this view, and so one might conjecture that the 'good grounds' lie in another direction; but, if so, there is no indication where. As for the difficulties with Carnap's suggestion, it is reasonable to assume that they are revealed in his unpublished "Is mathematical syntax of language?"; and I will discuss them in the next section.

3 'Platonism'

In any case, Gödel concludes in [*1933o] that

[19] Providing the rank of ϕ is a 'predicative ordinal' in the sense of [?] and [Schütte, 1965], transfinite induction up to this rank is deducible in predicative analysis involving only formulas of lower rank; and so there is no circle here.

> ... our axioms, if interpreted as meaningful statements, necessarily presuppose a kind of Platonism, which cannot satisfy any critical mind and which does not even produce the conviction that they are consistent. (p. 50)

This refers back to his earlier comments on non-constructive existence ("just as if in some objective realm of ideas this question [of existence or not] were settled quite independently of any human knowledge") and impredicativity (" this process of definition presupposes that the totality of all properties exists somehow independently of our knowledge and our definitions, and that our definitions merely serve to pick out certain of these previously existing properties.") The remark is surprising in view of Gödel's avowed Platonism, both in his published papers [1944; 1964] and in some of his unpublished work. For example, in [*1951, pp. 322-23] he writes

> I am under the impression that after sufficient clarification of the concepts in question it will be possible to conduct these discussions with mathematical rigor and that the result then will be that (under certain assumptions which can hardly be denied [in particular the assumption that there exists at all something like mathematical knowledge]) the Platonistic view is the only one tenable. Thereby I mean the view that mathematics describes a non-sensual reality, which exists independently of the human mind and is only perceived, and probably perceived very incompletely, by the human mind.

One explanation given, for example in [Davis, 1998], for the conflict between these passages is that Gödel's philosophical position evolved from an earlier anti-Platonism to the explicit Platonism of [1944; 1947]. One difficulty with this view arises from a questionnaire sent to Gödel by B.D. Grandjean in 1974,[20] which Gödel partially filled out but never sent. He states in response to the question of whether he held the view "described by some as 'mathematical realism', whereby mathematical sets and theorems are regarded as describing *objects* of some kind" in the 1920's and early 1930's, that he held this view since 1925. In a typed but unsigned and unsent letter to Grandjean, he wrote "I was a mathematical and conceptual realist since about 1925." Davis was aware of this response, but seems to discount it in the face

[20] See [Wang, 1987, pp. 16-21] for a description of the questionnaire and of Gödel's response.

of the remark from [Gödel, *1933o] just quoted and other evidence that "It was some years before Gödel's own mind was satisfied by this very 'kind of Platonism' ". In particular, he refers to Gödel's expression of regret as late as 1938 over the failure of Hilbert's program, quoting from [Gödel, *1938a, 113] "If the original Hilbert program could have been carried out, that would have been without any doubt of enormous epistemological value." But one could believe that (as I do) and still be a Platonist (as, in one sense, I am). Moreover, this remark is in the context of summing up the progress made in the extended Hilbert program, and Gödel is comparing its achievements with what might have been (in some funny sense of "might have been"). Note that Gödel retained interest in constructive foundation for classical mathematics throughout his life: for example, he was thinking about the Dialectica interpretation in the 1970's, long after the explicit expression of Platonism in [1944]. One challenge for the evolutionary view of Davis is to explain what the evolution was in reaction to. Gödel's later arguments for Platonism were, as we shall see, based primarily on the failure of alternatives. But the data upon which he pronounced this failure, in particular his own incompleteness theorems, were known to him by 1933. Perhaps an answer is that in 1933 he still thought, as he did in 1931, that his incompleteness theorems did not necessarily undermine Hilbert's program (although by that time he did identify finitism with primitive recursive arithmetic); but then his expression of regret over its failure in 1938 cannot be taken as evidence of a continued anti-Platonism. I don't mean by this to say that Davis is wrong; but for me at least, there is still no entirely convincing explanation of the conflict between Gödel's different statements about his Platonism.

Some light may be shed on the historical question of Gödel's commitment to 'Platonism' and, at the same time, some reinforcement might be gained for the view that the terms "Platonism", "realism", "objectivism", etc., all lack the precision to be useful labels for philosophical positions, by distinguishing two quite different positions, one weaker than the other, involved in the discussion of Gödel's Platonism.[21] It is not clear that Gödel himself did distinguish between them; but, as we shall see, the arguments that he actually gives for Platonism support the weaker position but not the stronger.[22] The weaker sense of 'Platonism' or 'realism' is the view that terms in math-

[21] [Parsons, 1995b] contains an extensive and penetrating discussion of Gödel's Platonism and his concept of mathematical intuition.

[22] On the other hand, it could be the distinction between realism about objects and realism about concepts, which he mentions in the response to Grandjean. See below.

ematical propositions, such as propositions in number theory or set theory, denote *sui generis* objects, i.e. which are not physical objects, nor mental objects nor fictions, for example, and are not analyzable way. Gödel seems to be subscribing to ‘mathematical realism’ in the response to Grandjean’s questionnaire in at least this sense. Much of what is commonly ascribed to the Platonist point of view follows from Platonism in this sense. For example, it follows that mathematical objects such as numbers, functions and sets are *outside space and time* (or, as some writers express this, are *abstract*), are *independent of the human mind* and are *not created* (since otherwise they would be in time), etc.. It sometimes makes sense on this view to say that these objects are *discovered*, namely when, as in the case of Cantor’s transfinite numbers, there is a more or less definite act of creating their theory. But were the finite whole numbers discovered? If so when and by whom?[23]

In his introductory note (p. 303), Boolos writes concerning the above quote from [Gödel, *1951, pp. 322-23] “What is surprising here is not the commitment to Platonism, but the suggestion, which recalls Leibniz’s project for a universal characteristic, that there could be a mathematically rigorous discussion of these matters, of which the correctness of any such view could be a ‘result’.” But if we understand “Platonism” in the above sense, then I find neither the Platonistic position nor the view that it could be rigorously demonstrated surprising. For Platonism in this sense is the *default* position. Defending it on the assumption that the theorems of arithmetic and analysis, for example, are true (“that there exists at all something like mathematical knowledge”) requires only showing that there is no alternative way of construing arithmetical or analytic propositions, so that the apparent reference to numbers, sets and the like is analyzed away. In fact, in many cases, Gödel’s affirmations of Platonism have been in the context of criticizing such alternatives and so his argument in these contexts is simply an argument for Platonism in this default sense. In particular, this is so in [*1951], although he notes there that his discussion does not yet constitute a proof of Platonism:

> Of course I do not claim that the forgoing considerations amount to a real proof of this view about the nature of mathematics. The most I could assert would be to have disproved the nomi-

[23]There is a more extensive discussion in [Tait, 2001, §II] of the extent to which many Platonist claims about mathematics may be understood as remarks about the grammar of mathematical propositions, when these are taken at face-value.

> nalist view, which considers mathematics to consist solely in syntactical conventions and their consequences. Moreover, I have adduced some strong arguments against the more general view that mathematics is our own creation. There are, however, other alternatives to Platonism, in particular psychologism and Aristotelian realism. In order to establish Platonistic realism, these theories would have to be disproved one after the other, and then it would have to be shown that they exhaust all possibilities. I am not in a position to do this now; . . . [pp. 321-22]

It seems entirely plausible that Gödel is right that the alternatives to realism in this sense are refutable.

Another way to try to refute Platonism in the default sense, that is, other than attempting some kind of reduction, would be to provide a convincing and non-question-begging analysis of what it means to exist and then to show that mathematical objects don't exist in this sense. Given the extensive anti-Platonist literature, it is striking that no serious attempt has been made to establish such a criterion for existence; but in any case, it does not seem likely that a plausible one can be found. What is common to the notion of existence across different domains of discourse is purely formal, expressed by the logical rules of quantification.

From this point of view, the assertion that mathematical objects don't exist makes no sense: as a mathematical assertion, it has trivial counterexamples. As an assertion external to mathematics, it depends on a univocal notion of existence which is simply not forthcoming. So, in a sense, realism in the default sense is not a substantive philosophical position: if it is meaningless to deny the existence of mathematical objects in the external sense, then it is also meaningless to affirm their existence in this sense. It is only as a meta-theoretical stance that realism is substantive, defending the legitimacy of ordinary mathematical practice against scepticism.

For this reason, too, Platonism in this default sense does not require for its justification the elimination of alternatives—that it be *proved* in Gödel's sense. For example, the reduction of number theory to something else should not lead us to reject numbers as objects; indeed, it should just increase our confidence in the coherence of what we say about them.

Platonism in the default sense does not imply, as it is often assumed to, classical as opposed to constructive mathematics. Indeed, it does imply the rejection of certain conceptions of constructive mathematics, including Brouwer's, according to which mathematical objects are rejected as *sui generis* and are to be understood in terms of mental constructions. But this does not seem to be the position of many present-day constructivists who, whatever the validity of their grounds for restricting mathematics, do not really attempt to reduce the mathematical ontology to something else. (Indeed, if truth is replaced by existence of a proof, then proofs, too, are part of the ontology and are not to be analyzed away.)

One may feel that the first passage in [*1951] quoted above commits Gödel to Platonism in more than the default sense, because he speaks of 'perception' of mathematical reality. But even in sense perception there is a conceptual element: perceiving a physical object or a person, for example, means perceiving it *as* such. In the Supplement to [1964], Gödel is explicit about this: he writes

> That something besides the sensations actually is immediately given follows (independently of mathematics) from the fact that even our ideas referring to physical objects contain constituents qualitatively different from sensations or mere combinations of sensations, e.g. the idea of object itself, whereas, on the other hand, by our thinking we cannot create any qualitatively new elements, but only reproduce and combine those that are given. [p. 271-2]

Objects are 'given' to us in perception only against a conceptual background. In the case of the 'non-sensual reality' of mathematics, only the conceptual element is present and, in the sense that numbers or sets or the like are given to us, it is entirely conceptual. I am not anxious to defend the use of the word "perception" in this connection; but I don't see it as having in itself implications that go beyond Platonism in the default sense.

On the other hand, that passage is followed by

> It by no means follows, however, that the data of this second kind [the 'given' underlying mathematics], because they cannot be associated with actions of certain things on our sense organs, are something purely subjective, as Kant asserted. Rather they, too, may represent an aspect of objective reality, but, as opposed to sensations, their presence in us may be due to another kind of relationship between ourselves and reality. (W.T.)

He goes on to assert that "the question of the objective existence of the objects of mathematical intuition" is "an exact replica of the question of the objective existence of the outer world." This latter remark is of course compatible with realism in the default sense: one need only hold that both questions are, as Wittgenstein put it, merely questions about the natural form of expression and are devoid of real substance. But one begins to have the sense from the preceding quote that, for Gödel, there *is* a real question of objective existence—that the whole of arithmetic and analysis, for example, although internally consistent, might nevertheless be false because there does not happen to be anything in objective reality which supports their truth; and this clearly goes beyond the default sense of Platonism described above. The comparison of the physical and mathematical realms in the above passage somewhat echoes his earlier discussion in [1944]

> It seems to me that the assumption of such objects [viz. classes and concepts] is quite as legitimate as the assumption of physical bodies and that there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions and in both cases it is impossible to interpret the propositions one wants to assert about these entities as propositions about the "data", i.e., in the latter case the actually occurring sense perceptions. (W.T.)

Here, too, the evidence is mixed. The last clause can be understood as an argument for realism in the default sense. But the earlier part of the passage seems to imply that a belief in the existence of physical or mathematical objects in general is something that could be wrong in a non-trivial sense—e.g. that it could be the case, not that this or that physical object fails to exist, but that physical objects *in general* fail to exist. So, too, in the mathematical case, classes and concepts in general (and not just classes or concepts with this or that particular property) could fail to exist. Such a view goes beyond realism in the default sense; for it is compatible with the latter that the question of whether or not numbers, functions, sets, etc. in general exist, as a question external to mathematics and challenging its validity, simply has no real content.

As was noted in [Tait, 1986, note 3], the passage quoted above from [Gödel, 1964] and, in particular, the sentence "Rather they, too, may represent an aspect of objective reality, but, as opposed to sensations, their presence in us may be due to another kind of relationship between ourselves and reality" *is* clear evidence that Gödel's realism does go beyond the default sense.[24] Perhaps the most conclusive evidence for this is in [*1953/9] and [*1961/?]. The former consists of two versions, number III and a much more concise V, out of six of a paper entitled "Is mathematics the syntax of language?" that Gödel intended for the Schilpp volume on Carnap but in fact never published. In this paper, he argues against the view that mathematics is the syntax of language on the grounds such a view would entail a commitment to prove the rules of language in question to be consistent, which can't be done without invoking further mathematics, not encompassed by these rules. An excellent evaluation of this argument in relation to Carnap's own aims is contained in Goldfarb's introductory note to the paper. But the point that I want to make here is that Gödel regarded his realism as an antidote for this problem: the axioms of mathematics are true of the objective domain of mathematical objects, and so they cannot imply a contradiction. Now, for realism in the default sense, we may indeed talk about domains of mathematical objects, e.g. the system of natural numbers or of all pure sets, and we would certainly say that the axioms of arithmetic or set theory, respectively, are true in these systems. But we may also hold that these systems are constructed *within* mathematics. On this view, we cannot conclude that the axioms of mathematics are consistent because they hold in these structures: if the axioms are inconsistent, then our conception of these structures is just incoherent. Of course, *within* mathematics we *can* conclude consistency of some set of axioms, e.g. the axioms of ZF from the existence of a model. But the conclusion is based on just the axioms in question together with some new ones, e.g., the assumption that there is an inaccessible cardinal; and so this would not satisfy Gödel's quest for an *external* guarantee of consistency.

[24]Parsons [1995b, p.67, footnote 44] suggests that I rejected in that paper the interpretation of Gödel as an "archetypical Platonist". In my note I was concerned to correct a misreading of Gödel's Supplement; but I think that it is clear that I indicated that this very passage was evidence that Gödel's realism went beyond realism in the sense that I defended in that paper and which I am now referring to as the default sense. However, I would say that, in all of Gödel's *published* work (as opposed to the unpublished papers which we are now discussing), the passage in question is the only one which conclusively commits him to realism in a stronger sense.

When we ask how we are to know that our axioms are true of the objective domain of mathematical objects, Gödel's answer is that we have an intuitive grasp of truth in this domain. As we have sense perception in the case of empirical facts, we have intuition in the case of mathematical facts. As Parsons points out in [Parsons, 1995b], Gödel's notion of intuition here is not the Kantian notion which is reflected in Brouwer's use of the term or in Hilbert's use of it in, say, [Hilbert, 1926] when describing the finitist conception of mathematics: intuition for Gödel yields new axioms of set theory. Also, for the same reason, it differs from what we often express when we say something is intuitively true, meaning that we ought to be able to prove it—the intuition providing some indication of *how* to prove it. For Gödel intuition provides evidence of truth of propositions to be taken as axioms, i.e. as the *basis* of proof. The role that mathematical intuition plays for him is clear from [*1961/?, p.385], where he writes

> I would like to point out that this intuitive grasping of ever newer axioms that are logically independent from the earlier ones, which is necessary for the solvability of all problems even within a very limited range, agrees in principle with the Kantian conception of mathematics. The relevant utterances by Kant are, it is true, incorrect if taken literally, since Kant asserts that in the derivation of geometrical theorems we always need new geometrical intuitions, and that therefore a purely logical derivation from a finite set of axioms is impossible. This is demonstrably false. However, if we replace the term "geometrical" by "mathematical" or "set theoretical", then it becomes a demonstrably true proposition.

In the case of set theory, what is 'demonstrably true' is that, whatever axioms S we have accepted as true of the universe of sets, there will be propositions undecided by these axioms. When the engine producing these propositions is Gödel's incompleteness theorems, so that the proposition undecided by S is for example $Consis(S)$, then it seems right that we are led immediately to accept this proposition; and we can call this an act of intuition. But when the engine is Lévy's reflection principle or some extension of it, then, as I argued in §1, there may well be some difference of opinion about accepting the undecidable proposition produced. So it is not clear that this more powerful engine for producing undecidable sentences really does lead us to new axioms on which we would have even consensus agreement. Kant grossly overestimated the amount of hard-wiring in the common human mind in connection with geometry; it seems possible that Gödel commits the same error in connection with set theory.

Nevertheless, Gödel is committed to the view that intuition is a means of *discovering* new truths, where these truths are somehow already implicit in our concepts, but not logically derivable from the axioms we have already accepted concerning these concepts. This would seem to be an instance of Gödel's 'conceptual realism'. There is a discussion of it in [Wang, 1974, pp. 84-85].[25] I quote Wang:

> 'If we begin with a vague intuitive concept, how can we find a sharp concept to correspond to it faithfully?' The answer Gödel gives is that the sharp concept is there all along, only we did not perceive it clearly at first.

The context of this quote is the discussion of Turing's analysis of the notion of mechanical computation; but the view expressed seems quite general. In particular, applied to the present case, the concept of ordinal number is already determined in the sense that it is already determined what totalities of ordinals are bounded.[26] It follows that new axioms that we introduce to give upper bounds to such totalities are already contained in that concept. So realism about concepts, so understood, claims that mathematical truth is prior to the adoption of axioms rather than founded on them. If this is the right way to understand Gödel, then in recognizing the distinction between realism about mathematics (i.e. mathematical objects) and realism about concepts, he in fact did make our distinction between realism in the default sense and realism in the wider sense. Also, understood this way, the distinction between Gödel's realism about mathematics and his realism about concepts would seem to coincide with Parson's distinction between *object Platonism* and *truth Platonism* [Parsons, 1995a], although it is not clear to me that Parsons himself in [1995b] or [n.d.] identifies Gödel's realism about concepts with truth Platonism. On the other hand, Parsons writes "Although I can't defend this claim here, I don't think Gödel has as good an argument for the claim that the assumption of concepts on his conception of them is 'necessary for a satisfactory system of mathematics' as he has for

[25] In his response to the Grandjean questionnaire, Gödel explicitly cites this passage as an expression of his views.

[26] [Parsons, n.d.] quotes from Wang's reconstruction of his conversations with Gödel: "To say that the universe of all sets is an unfinished totality does not mean objective indeterminateness, but merely a subjective inability to finish it." [Wang, 1996, 8.3.4].

the corresponding more straightforward claim about sets" [n.d.]. This corresponds very well to the situation as we see the distinction: the argument for the default position, although not completely given, is that no alternative preserves mathematical knowledge. No argument is really given for the stronger position, that there is a transcendental notion of truth to which our axioms must conform. Indeed, there is a strong argument against this view: if the new axioms we might introduce could be false, then that should be the case also for the axioms that we have already introduced.[27] Thus, the axioms do not *define* the subject matter; rather, they are merely an expression of our grasp of it and could be false. It follows that mathematics is a speculative science and Gödel's comparison of it with physical science takes on a sharper—and to my mind, unfortunate— meaning, and perhaps his second criterion for accepting new axioms [1964, 265], mentioned above in §1, makes better sense.

The view that the axioms do define the subject matter is perfectly compatible with realism in the default sense, in spite of the fact that it is often identified with formalism or with the view that 'mathematics is the syntax of language'. It is the conception of mathematics of Hilbert: we establish a system of objects when we specify, in the form of an axiomatic theory, how to reason about them. Of course, Hilbert demanded of his axiom system that it be complete and consistent and we know now both that no system S_α is complete and that there is no elementary proof of its consistency. But, in spite of Gödel's attack on Carnap in "Is mathematics the syntax of language?" because of the impossibility of an elementary consistency proof, it seems to me that the axiomatic point of view remains viable and, indeed, is the only viable one. I won't discuss this further here, but[Tait, 2001] contains an extensive discussion of it.

4 Extensions of Finitism

Having criticized the non-constructive and impredicative methods expressed by the axioms of set theory in [*1933o], "which, if interpreted as meaningful statements, necessarily presuppose a kind of Platonism, which ... does not

[27]Here I am ignoring the special problem that the new—or old—axioms could turn out to be inconsistent. If this should happen, then on one view, it would entail a more or less serious revision of our concepts. But on Gödel's view, apparently, it would only mean that we falsely analyzed the concepts.

even produce the conviction that they are consistent", Gödel turns to the question of giving consistency proofs for the "objectionable methods", where they are regarded purely formally and without regard for their meaning—in other words, he turns to Hilbert's program. He writes "Of course, the chief point in the desired proof of freedom from contradiction is that it must be conducted by perfectly unobjectionable methods ..." and goes on to identify constructive mathematics as what remains when we drop the objectionable methods. He notes that there are different layers of constructive mathematics, beginning with the "strictest form of constructive mathematics", which he refers to as the "system A", and converging towards non-constructive mathematics. The system A appears to be just primitive recursive arithmetic PRA, and in both [*1933o, p. 52] and [*1938a, p. 93], he explicitly identifies it with Hilbert's finitism.[28] Later, in [1958], he distinguishes this lowest level of finitism from its extensions:

> Since finitary mathematics is defined as the mathematics of *concrete intuition*, this seems to imply that *abstract concepts* are needed for the proof of consistency of number theory. ... By abstract concepts, in this context, are meant concepts which are essentially of second or higher level, i.e., which do not have as their content properties or relations of *concrete objects* (such as combinations of symbols), but rather of *thought structures* or *thought contents* (e.g. proofs, meaningful propositions, and so on), where in the proofs of propositions about these mental objects insights are needed which are not derived from a reflection upon the combinatorial (space-time) properties of the symbols

[28]In [*1938a, p. 105], moreover, he identifies ω^ω as the 'ordinal' of "finititary number theory", i.e. as the bound on the ordinals α such that definition of functions by recursion on α can be finitistically justified. ω^ω is in fact the ordinal of PRA. It was known at the time that recursion and (suitably formulated) induction up to any ω^n was derivable in PRA; but, as the editors point out (p. 79), it was only in the early 1960's that a detailed proof was available that recursion and induction up to ω^ω is not derivable in PRA. On the other hand, it was well-known in the 1930's (e.g. see [Hilbert and Bernays, 1934, §7] that the enumerating function (of two variables) of the primitive recursive functions of one variable, which is clearly not primitive recursive, can be defined by a double nested recursion (i.e. a nested ω^2-recursion). But it is rather easy to see, by considering its computations, that this function is definable by recursion on ω^ω. Moreover, using methods utilized in [Hilbert and Bernays, 1934], the consistency of PRA can be proved in PRA together with this enumerating function.

> representing them, but rather from a reflection upon the *meanings* involved.[29]

But I don't like this characterization of the difference between finitism and its extensions: speaking from the propositions-as-types point of view, functions and proofs are simply objects of higher types than the type $\mathbb{N}$ of the numbers; and this is a logical, not an epistemological difference (see [Tait, 1981]). Perhaps Gödel sees the difference in the fact that we recognize a number sign, a numeral, purely in terms of its syntactical structure, whereas this is not so for functions and proofs. Thus, we may be given some defining equations or a Turing machine for a numerical function, but determining whether or not it really defines a (total) function is not a matter of inspecting it syntactically: we must appeal to its meaning. But one may take the position in constructive mathematics that functions of a given type and proofs of a given proposition, too, are given by definite syntactical rules of construction, analogous to the rules $\Rightarrow 0$, $n \Rightarrow S(n)$ for constructing numbers, such that objects constructed by such rules are immediately recognized for what they are. That we may define a function by means of a Turing machine, where there is then a burden of proving that a function is actually determined (and where the proof itself amount to representing it in the canonical notation), is not different from defining a number by means of some property, with the additional burden of extracting from that definition a numeral. There remains only this difference between the case of numbers and higher type objects: the above rules for constructing numbers are given once and for all. In the case of functions or proofs, the rules are (in analogy with the axioms of set theory) essentially incomplete: the addition of new types can (using, for example, impredicative primitive recursion) lead to the construction of new functions or proofs of a given type (e.g. where that type may be a consistency statement for some formal system).

Although Gödel identified Hilbert's finitism with PRA in [*1933o; *1938a], he writes immediately following the passage just quoted that "Due to the lack of a precise definition of either concrete or abstract evidence there exists, today, no rigorous proof for the insufficiency (even for the the consistency proof of number theory) of finitary mathematics." Moreover, as we have already noted, he explicitly denies in his paper on incompleteness that his results undermine Hilbert's program to obtain *finitary* consistency proofs. Of course, one explanation for the discrepancy might be that he recognized

[29] I have actually taken this quote from the slightly revised text of [1972, pp. 272-3].

the difference between the conceptual question of how "finitism" should be understood and the historical question of what Hilbert meant; and, while he felt that PRA has a special claim to be called "finitism", he was leaving it open whether or not that is the sense of finitism that Hilbert had in mind.[30]

In [*1933o] the only extension of finitism that he discusses, and that negatively, is Brouwer's intuitionism and its formalization by Heyting; but he ends with the expression of hope that "in the future one may find other and more satisfactory methods of construction beyond the limits of the system A." In [*1938a] he again takes up the question. But by that time, other possibilities for extending finitism had emerged: he explicitly mentions three extensions: (i) The use of functions of higher type. This idea is developed in some detail in [*1941] and then later in [1958]. (ii) The use of modal logic. Here Gödel is again referring to intuitionism and in particular to Heyting's interpretation of the logical operations in terms of the notion 'it is provable that…'. (iii) The use of transfinite induction and recursion. Gentzen [1936] had by that time given a proof of the consistency of Peano Arithmetic PA which is formalizable in PRA extended by a quantifier-free version of the principle of induction on an ordering of the natural numbers of type ϵ_0.

In [*1938a, p. 91], Gödel lays down four requirements which he regards as necessary, if not sufficient, for a formal system to be constructive. The first is that the primitive operations and relations should be computable. The second is that $\exists$ should not occur among the primitive signs (the assertion of $\exists x F(x)$ being introduced as a shorthand for having a proof of $F(t)$ for some particular term t) and $\forall$ should not occur in the scope of the propositional connectives. The third concerns axioms and rules of inference, which may include those of A, but he leaves it open that there may be others. The last requirement is that "objects should be surveyable (that is, denumerable)". The first three requirements would seem to be closely connected: presumably, the computation of the primitive operations and the deciding of the primitive relations will be according to the defining axioms for these objects. And the second requirement, concerning the quantifiers, may be seen as a consequence of a more far-reaching requirement that, not only the primitive, but also the definable operations and relations should be computable and decidable. These requirements seem less than compelling, even from a constructive point of view. Although, certainly since Kronecker, it has been a part of the

[30]For a further discussion of the conceptual question, see [Tait, 1981] and for discussions of the historical question, see in addition [Tait, n.d.; Zach, 1998].

finitist creed that all concepts must be decidable, there is after all room for the more expansive constructivism that admits undecidable concepts, but rejects applying the law of excluded middle to them. This already throws the first three requirements into question. It will be useful to discuss the fourth requirement after a discussion, in the next section, of Gödel's contention that his theory of functions of higher type satisfies it.

Speaking personally, reading [*1938a] was something of a shock. In 1958-9, at the beginning of my teaching career, I attended a course of lectures by G. Kreisel at Stanford.[31] I recently looked through Kreisel's thermofaxed hand-written notes for the lectures and realized that they could have been an extended commentary on [*1938a], containing essentially no new ideas and reflecting remarkably little progress in the intervening twenty years. [*1938a] consists of notes, sometimes rather cryptic, for a lecture. The editors, Charles Parsons and Wilfried Sieg, making use of an earlier draft of notes for the lecture, have done a good job of piecing together a coherent text, as well as producing in the editors' introduction a scholarly analysis of the text and its historical context.

5 The Dialectica Interpretation

The main result in direction (i), Gödel's so-called *Dialectica* interpretation, which was first published in [1958], is already essentially contained in [*1941], the text of a lecture delivered at Yale. In his introductory note, Troelstra refers to a lecture at Princeton in the same year on intuitionistic logic and the Dialectica interpretation. A later version of the *Dialectica* paper, [Gödel, 1972], was not published by Gödel but appears in [Gödel, 1990].

The interpretation associates with each formula ϕ of first order intuitionistic arithmetic HA a formula

$$(1) \qquad \exists u \cdots v \forall x \cdots y F(u, \ldots, v, x, \ldots, y, \vec{z})$$

[31] I don't remember how many students started out in the course; but after a week or so there were two students, and then one student, in addition to me, meeting finally in Kreisel's closet-cum-office. I don't really remember the surviving students either, but that they lasted so long was, I think, due simply to stubbornness: I don't believe that they understood much of the lectures at all. As for me—I won't speak for Kreisel—I doubt that I really understood more than seventy percent of the material either, although the course and my discussions with Kreisel at that time were invaluable to me.

where each of the variables $u, \ldots, v, x, \ldots, y$ ranges over some finite type, $\vec{z}$ is a list of the free variables in ϕ, and $F(u, \ldots, v, x, \ldots, y, \vec{z})$ is quantifier-free. Gödel's result is that, from a deduction of ϕ in HA, we may read off constant terms $s, \ldots, t$ of suitable type and a deduction of formula

$$F(s, \ldots, t, x, \ldots, y, \vec{z})$$

in a certain quantifier-free theory T of the *impredicative primitive recursive (ipr) functions of finite type*, where the atomic formulas are equations between terms of the same type.[32] The *ipr* functions are closely related to Hilbert's hierarchy [1926], except that Hilbert's hierarchy extends through the second number class. Both admit precisely those functions built up from 0 and the successor function by means of explicit definition and primitive recursive definition. We speak of the *impredicative* primitive recursive functions, because, as we noted in §2, one may define functions Φ of some type by primitive recursion which can then be used to define functions of lower type and, indeed, arguments for Φ.[33] If ϕ is quantifier-free, then $F(s, \ldots, t, x, \ldots, y, \vec{z})$ is just ϕ, and so the interpretation yields a consistency proof for HA relative to T. This extends to a consistency proof for classical first-order arithmetic PA, since, as Gödel had already shown in [1933e], PA can be regarded as a subsystem of HA by eliminating $\vee$ and $\exists$ in favor of their de Morgan equivalents. In [*1941, p. 200], he suggests the possibility of extending his result to analysis (second-order number theory) by using transfinite types. This has never been realized. [Spector, 1962] extended the interpretation to analysis by adjoining to T the clearly non-constructive principle of *bar recursion* of arbitrary finite type; but the types remain finite. An excellent survey of the

[32][*1941] is not explicit about what types are admitted for these equations. In [1958], equations of arbitrary type are admitted, where the equality of objects of higher type is understood to be *intensional* or *definitional equality.* Troelstra states that, in the Princeton lecture, Gödel explicitly restricts the equations to numerical terms. As far as the interpretation of HA in T is concerned, however, it doesn't matter whether or not equations of higher type are admitted.

[33]Thus the hierarchy of *ipr* functions is distinct from the hierarchy of *predicative* primitive recursive functions of finite type, introduced in [Kleene, 1959]. It had been noted already in [Hilbert, 1926] that the Ackermann function, which is not Kleene primitive recursive, is *ipr*. Actually, Hilbert indicates that his types are those built up from $\mathbb{N}$ by means of the operation $A \mapsto A \longrightarrow \mathbb{N}$; but it is clear from his examples that he intends functions of more than one variable or, what amounts to the same thing, the types built up from $\mathbb{N}$ by means of the operation $\longrightarrow$.

Dialectica interpretation and its subsequent development can be found in [Avigad and Feferman, 1998].

Concerning the four requirements that constructive formal systems should satisfy, Gödel states in [*1938a, p. 94] that, among the three extensions of PRA that he listed, only the extension obtained by adding functions of finite type, i.e. presumably the system T, meets all four requirements. (In [*1941] he mentions only the first two of the requirements on constructive systems.) But there are problems with this assessment, concerning the first and fourth requirements.[34]

Taking the fourth requirement first, in what sense are the objects of finite types other than $\mathbb{N}$ surveyable, i.e. denumerable? Gödel, himself, remarks (p. 97) that the fourth requirement is problematic because of the notion of function. About the first requirement, he says that it is not 'sharp'. Both in [*1938a] and in [1958], Gödel justifies the definitional axioms of T on the grounds that they define computable functions. In the later paper, he gives the following definition: the computable functions of type $\mathbb{N}$ are the natural numbers. A computable function of type $A \longrightarrow B$ is an "operation, always performable (and constructively recognized as such), that to every [function of type A] assigns a function of type [B]".[35] Presumably what is not sharp is the notion of an 'operation, always performable (and constructively recognized as such)'. Thus, the totality of functions of a given type is denumerable, because each function is given by a rule of computation, but this is problematic because we cannot give, once and for all, a list of definitions which will yield all and only computable functions of that type. So the editors must be correct in supposing that, when Gödel states that only the first extension, namely by functions of finite types, satisfies all the requirements, he has in mind the relativization of the notion of a computable function of finite type to some restricted class of definitions, containing at least those formalized in T itself—for example, we can relativize it to precisely those definitions formalized in T. Now the objects of T are indeed surveyable, in the sense that we can effectively enumerate the closed terms of T of any given type. But of course this enumeration, as an enumeration of the objects, will involve repetitions. In order to eliminate these, we would need to know when two terms denote the same object, i.e. when they are definitionally equal. In

[34]See the discussion of this in the introductory note on pp. 69-70.

[35]The paraphrases are because Gödel considered types of functions of several variables, whereas I am speaking only of functions of one variable, which in the presence of higher types is sufficient, as essentially noted in [Schönfinkel, 1924].

fact, in [1958], the formulas of T are built up from equations between objects of arbitrary finite type; so in this case, the first requirement already requires that this relation be decidable.

The question of the decidability of definitional equality is closely related to that of Gödel's first requirement for constructivity. We noted that in [1958] Gödel defined a computable function of type $A \longrightarrow B$ to be an "operation, always performable (and constructively recognized as such), that to every [computable function of type A] assigns a computable function of type [B]". But surely the relative virtue of a system such as T will depend on what constructive methods are needed to see that the operations involved are always performable. [*1941, 195] waives this question after noting simply that, if f is a computable function of type $\mathbb{N} \longrightarrow [A \longrightarrow A]$ and b is an object of type A (i.e. a numeral if $A = \mathbb{N}$ and a computable function of type A otherwise), and if g of type $\mathbb{N} \longrightarrow A$ is defined by primitive recursion: $g0 = b, g(x+1) = fx(gx)$, then it immediately follows from the definition of computability that go, and so $g1$, and so $g2$, etc., are all defined and computable.[36] Although he writes "A closer examination of the question in which manner the functions obtained by these two schemes [i.e., explicit and primitive recursive definition] are really computable is pretty complicated," in [1958, p. 283, footnote 5] he emphasizes the fact that we must *immediately* see that these operations are "always performable"; otherwise, the notion of computability will depend upon the notion of proof, and it is the latter notion that he is attempting to dispense with in giving his interpretation of HA. But it is hard to see how proof is to be avoided. Gödel did not, in the above example, appeal to immediate insight that primitive recursive definition is justified: he gave the argument, as he must. And the argument is by mathematical induction, concluding that gz is computable for all numbers z, from the premises $g0$ is computable and, if gz is computable, then so is $g(z+1)$. Moreover, depending on the type A of gn, this property of being computable is logically quite complex. Thus, in terms of the term model, Gödel's definition of the computable functions takes the form

$$C_{\mathbb{N}}(t) := \ t \text{ is a closed term definitionally equal to a numeral}$$

$$C_{A \longrightarrow B}(t) := \forall x[C_A(x) \longrightarrow C_B(tx)]^{37}$$

[36] Gödel's example is actually only the special case $g0 = b, g(z+1) = f(gz)$ of iteration.

[37] If we replace the definition of $C_{\mathbb{N}}(t)$ by the condition that t is a *well-founded* closed

So the proof that the above function g is computable is by mathematical induction applied to the predicate $C_A(gz)$, which, coding the terms of T by their Gödel-numbers, we may regard as an arithmetical predicate. Now define the *level* of the type A to be $l(A) = Max\{l(A_i)+1 \mid i < n\}$, where A is uniquely of the form $A_1 \longrightarrow (\ldots \longrightarrow (A_{n-1} \longrightarrow \mathbb{N})\ldots)$. Let

$$\exists u \cdots v \forall x \cdots y F(u, \ldots, v, x, \ldots, y, z)$$

be the *Dialectica* interpretation of $C_A(gz)$. Then the level of A is precisely the maximum of the levels of the types of $u, \ldots, v$, say it is the level of the type B of one of these variables. Using the *Dialectica* interpretation to justify mathematical induction applied to $C_A(gz)$ requires defining a function h of type $\mathbb{N} \longrightarrow B$ by primitive recursion, and so we are back to a need to prove $C_B(gz)$ for all z. Thus, to justify primitive recursive definition at a given level, we need to apply mathematical induction to a certain formula, and to justify mathematical induction via the *Dialectica* interpretation applied to that formula, we need primitive recursive definition at that same level. As we mentioned above in §2, there is a circle here, too.

In a sense, Gödel was certainly aware of this circularity: in the introductory note to [*1941] Troelstra quotes from the Princeton lecture, where Gödel indicates that $C_A(t)$ can be proved in HA for all types A and closed terms t of type A.[38] There, Gödel rightly dismisses the argument as circular; but I believe that "a closer examination" of the question of computability has to lead precisely to this circle. In later years, Gödel was interested in the

term of type $\mathbb{N}$, meaning that every sequence $t = t_0, t_1, t_2, \ldots$, where t_{i+1} is obtained from t_i by replacing some definiens by the corresponding definiendum, is finite, then (i) for closed terms of type $\mathbb{N}$, well-foundedness implies definitional equality with a numeral, and (ii) the property $C_A(t)$, for any type A, implies that t is well-founded. Well-foundedness of every term implies that every term reduces to an irreducible term, which is unique by the Church-Rosser Theorem. So then definitional equality is decidable: reduce two terms to their unique irreducible form. They are definitionally equal just in case these irreducible terms are identical. With this definition of computability, $C_A(t)$ is deducible in HA for every term t of T of type A. With the original definition, this is true for all closed terms.

[38]This is part of the content of [Tait, 1967], a widely circulated version of which is in [Tait, 1963]. Troelstra's remark on this is too modest on behalf of Gödel: "This is strongly reminiscent of the computability method of Tait (1967)." I would rather say of Gödel: been there, done that. There is evidence that Gödel already knew this result in 1938.: once one knows that the computability of all terms of T can be deduced in HA, one easily obtains the interpretation of T in HA given in [Tait, 1967]. On p. 97 of [*1938a], he writes "With finite types one cannot prove the consistency of number theory. Very likely, this remark is based on the knowledge of that interpretation.

problem of assigning ordinals less than ϵ_0 to closed terms of T such that, if the term s reduces to the term t, then the ordinal of t is less than that of s. He speaks about the existence of such an assignment in the passage from the Princeton lecture quoted by Troelstra, but it seems clear that he never actually found one.[39] [?] solved this problem for a special class of reductions (but such that every reducible term can be reduced in the special way). But in any case, given such an assignment, the proof that every closed term of type $\mathbb{N}$ reduces to a numeral requires PRA together with transfinite induction up to ϵ_0, and so the argument that this extension of finitism satisfies even the first requirement of a constructive system would already require Gentzen's extension. Charles Parsons has pointed out to me that, although Gödel, in a draft of a letter Frederick Sawyer in 1974, continued to hold that T is more constructive than intuitionistic arithmetic (see [Gödel, 1990, p. 236, note k]), the letter was never sent. I agree entirely with Parsons' suggestion that this, together with his continuing work on the 1972 version of the *Dialectica* paper, might indicate that Gödel was not entirely comfortable with the position expressed in [*1938a; *1941] and in the letter to Sawyer.

6 Modal Logic/ Intuitionism

To some extent, perhaps, Gödel's negative view of intuitionistic logic rests on an overestimation of the scope of his interpretation of classical logic in intuitionistic logic using de Morgan equivalents (and double negating atomic formulas). Thus in [*1941] he writes "So intuitionistic logic, as far as the calculus of propositions and of quantifiers is concerned, turns out to be rather a renaming and reinterpretation than a radical change of classical logic." (p. 3) But even though the interpretation extends to the predicate calculus of finite type and as the editors' note, to a "(carefully formulated) set theory" (p. 72), it is not really the case that classical logic can be embedded in intuitionistic logic—at least for those of us who regard the axiom of choice in the second-order form

$$\forall x \exists y F(x, y) \longrightarrow \exists g \forall x F(x, gx)$$

[39]For example, my own last contact with him was a phone conversation in 1974 in which he expressed the desire to have this problem solved.

as a law of mathematical logic. For, although this is itself perfectly valid intuitionistically, its interpretation

$$\forall x \neg \forall y \neg F'(x,y) \longrightarrow \neg \forall g \neg \forall x F'(xgx)$$

where F' denotes the interpretation of F, is invalid. So the interpretation of the classical second-order predicate calculus in the corresponding intuitionistic system fails as soon as we add this principle.

But, Gödel's rejection of intuitionistic logic as a legitimate extension of the constructive point of view does not really depend on the his interpretation of classical logic in it: he takes the latter as merely a symptom of it. He objects to it on the grounds that it fails to satisfy all but the first of his requirements on a constructive theory. We have already discussed grounds for rejecting the first three of there requirements. As to the fourth, he seems to me to be invoking a double standard in relation to his theory T. Not all functions of finite type are defined in T: he explicitly notes that new ones are obtained by going to transfinite types. So his theory satisfies the fourth requirement only when we draw a line on what particular methods of defining functions we will admit. I think that his charge that intuitionism fails to satisfy the fourth requirement is based on a refusal to allow it the same freedom.

The real grounds for his rejection of intuitionism are revealed by his description of intuitionism as "the modal-logical route", where the modal operator in question is of course the *provability* operator $\mathcal{B}$, with $\mathcal{B}A$ meaning that A is provable. Again in [*1941] he writes

> the primitive terms of intuitionistic logic lack the complete perspicuity and clarity which should be required for the primitive terms of an intuitionistic system. E.g., $P \longrightarrow Q$ in intuitionistic logic means that Q can be derived from P, and $\neg P$ means that a contradiction can be derived from P. But the term "derived" cannot be understood in the sense of "derived in a definite formal system". (For this notion the axioms of intuitionistic logic would not hold.) So the notion of derivation or of proof must be taken in its intuitive meaning as something directly given by intuition, without any further explanation being necessary. This notion of an intuitionistically correct proof or constructive proof lacks the desirable precision. In fact one may say that it furnishes itself a counterexample against its own admissibility, insofar as it is

doubtful whether a proof utilizing this notion of a constructive proof is constructive or not.

The most reasonable interpretation of his parenthetical remark that "derived" cannot mean "derived in a definite formal system" is obtained from his paper "An interpretation of the intuitionistic propositional logic" [Gödel, 1933f] in which, reading the necessity operator $\mathcal{B}$ as meaning "it is provable that", he gives a translation of Heyting's propositional calculus into the modal system $S4$ such that every theorem of Heyting's system translates into a theorem of $S4$ and, for example, the translation of $p \vee \neg p$ is not a theorem.

$$\mathcal{B}[\mathcal{B}p \longrightarrow p]$$

is a theorem of $S4$, and so when he writes that $\mathcal{B}$ cannot be understood as provability in a formal system, he is referring to the fact that, if the formal system is consistent and includes HA, say, then the above theorem is false when we take p to be the formula $0 = 1$.

Heyting himself had introduced $\mathcal{B}A$ in a 1931 paper [Heyting, 1930b] as a meaningful proposition which, at least in some cases, is distinct from A itself. He gives as an example a universal arithmetical proposition $A = \forall x F(x)$. In this case $\mathcal{B}A$ asserts the existence of a construction, a proof, whereas A itself does not. He goes on to say that, in those cases in which A itself requires a construction, its meaning coincides with that of $\mathcal{B}A$. His example of this is any negation $\neg B$; for he takes the meaning of this to be that "B can be reduced to a contradiction". But in his later, more mature discussions of intuitionistic logic, Heyting asserts that *every* intuitionistic proposition A requires a construction: the assertion of A is warranted only on the basis of a proof, which is a construction of a certain type. So, on his later conception, he gave up the view that A and $\mathcal{B}A$ can be distinct propositions. Presumably he did do because, if A requires a construction, then what *other* construction could $\mathcal{B}A$ possibly require? The notion of proof is no more an ingredient of intuitionistic propositions than is the notion of truth an ingredient of propositions of classical mathematics.

In the lecture at Zilsel's, noting the non-constructive character of $S4$ as he interpreted it, Gödel goes on to say that this non-constructivity can be avoided by replacing provability as the basic notion by the proof *relation* '$z \vdash A$' or '$z \vdash A, B'$, meaning that z is a proof of B from A. However, he did not ' really develop the treatment of intuitionism in terms of the provability relation, and in particular he neither specifies the range of the variable z in

"$z \vdash A$" nor tries to define the meanings of the logical constants in terms of the provability relation. But later on, in Kreisel's 1958-59 lectures mentioned above in §4, this idea looms large: z is to range over all constructions and so what is needed for a foundation of intuitionism is a general logic-free theory of constructions. Thus, we are to begin with the primitive notion of a construction and, in terms of this, define the logical constants in such a way that intuitionistic theorems are valid. But, although Kreisel himself [1962; 1965], Nicolas Goodman, e.g., [1970], and others obtained limited results in this direction, in the sense that, given some particular formal intuitionistic system such as PA, one can give a theory of constructions that is adequate for it, the program in general was abandoned. One problem, even in the restricted cases in which it has been carried out, is that it seems necessary to assume that there is a decidable notion of proof of a Π^0_2 sentence. The argument for such a notion seems to be that, if one sees clearly that an intuitive argument for the sentence is correct, then it is a proof; if one does not see this clearly, then it is not. Perhaps not many people would be satisfied with that idea.

So the idea of starting with a given domain of constructions and then trying to define what it means for a particular construction to be a proof of a given proposition seems to have been abandoned. On the other hand, the foundations of intuitionistic logic do not require this idea. All that is needed is a specification of what counts as a construction or proof of a particular proposition A. And this differs in no essential way from a foundation for Gödel's theory T, where what is required is that, for any given finite type A, we know what it is to be an object of type A. The foundation of HA no more needs a general notion of construction than the foundation of T does. Indeed, as the Curry-Howard theory of propositions-as-types [Howard, 1980] makes clear, the formulas of HA constitute a natural generalization of the finite types of T, especially if we add to the latter the new type-forming operation $A \wedge B$ of forming pairs of objects of types A and B, respectively. (Gödel himself suggests this extension in [*1941, p. 196] as a means of simplifying the details of his interpretation.) In fact the natural rules for constructing objects of this type (pairing and projections) are precisely the rules of inference for conjunction $A \wedge B$, just as the rules for $A \longrightarrow B$ in T (λ-abstraction and evaluation of a function) are precisely the rules ($\longrightarrow$-introduction and -elimination) of implication. Moreover, the type-forming operations $N \longrightarrow B$ and $N \wedge B$ generalize to $\forall x B(x)$, the type of all functions of a numerical argument whose value for n is of type $B(n)$, and $\exists x B(x)$, the type of all pairs (n, b) where b is of type $B(n)$, respectively. The natural rules for constructing objects of these types again are precisely the rules of quantification and induction in HA. (To complete the description of the sentences of HA as types, $\perp$ and all false equations denote the null type and all true equations denote the one-element type.)

Of course, there is a certain incompleteness in this conception: by the introduction of new propositions/types into a particular system of types, say HA, with their corresponding introduction and elimination rules, we shall new proofs of propositions in the original system, in some cases of propositions which are unprovable in the original system. But this is the precise analogue of the incompleteness of the theory T, which Gödel explicitly noted: by going to higher types, we obtain functions of finite type not definable in T.

Viewed in this way, it is difficult to see why one should regard intuitionistic arithmetic as less constructive than T. Just as T is a theory of propositional combinations of equations between terms of the same finite type, so we can describe an extension of it, T^*, which is a theory of propositional combinations of equations between types and between terms of the same type in our extended sense of type. If it is the case that, for the intuitionist, the content of a proposition consists in what counts as a proof of it, then it is T^* that formalizes the content of intuitionistic arithmetic. I do not see where one could draw a principled line on the basis of which to call T more constructive than T^*.

7 Gentzen's Proof Theory

The last sections of [Gödel, *1938a] concern Gentzen's first version of his consistency proof for arithmetic [1936], which is founded on his notion of a *reduction*. We may express his idea in terms of a game $\mathcal{G}(\Gamma)$, which, unlike the game $\mathcal{T}(\phi)$ of §2, is played with finite sets of formal sentences of arithmetic rather than a single sentence. The game starts with the non-empty Γ on the board. Let Δ be on the board at a given stage. If it contains a true atomic sentence or a sentence together with its negation, then $\bigvee$ wins. If it consists entirely of false atomic sentences, then $\bigwedge$ wins. Assume neither is true. Then the game continues for at least one more stage. If Δ contains any $\bigwedge$ sentences, then $\bigwedge$ moves and replaces Δ on the board by the result of replacing each $\bigwedge$-sentence in Δ by one of its components. If there are no

$\bigwedge$-sentences in Δ, then it contains a $\bigvee$-sentence and $\bigvee$ moves by adding to Δ a component of one of the $\bigvee$-sentences in it, say ψ. (Gentzen allows that $\bigvee$ be given the choice also of dropping ψ; but that is an inessential difference. What *is* essential—and this is what distinguishes $\mathcal{G}(\{\phi\})$ from $\mathcal{T}(\phi)$—is that $\bigvee$ is allowed to *keep* ψ, so that if he makes a bad choice of a component the first time, he has a chance later on to choose a different one.) Gentzen's consistency proof consists in constructing, from each formal deduction of a sentence ϕ in PA, a winning strategy for $\bigvee$ in the game $\mathcal{G}(\{\phi\})$. Moreover, such a strategy may be given as a well-founded tree, whose paths consist of all the games played according to the strategy. In his proof, he, in effect, assigns to each such tree an ordinal $< \epsilon_0$ which is a bound on its height. [Schütte, 1951], at the cost of slightly obscuring the constructive content of Gentzen's result, recasts the construction as a cut-elimination, in which the cut-free deduction obtained from the given deduction is precisely the winning strategy.[40]

Gödel describes Gentzen's proof in [*1938a] in a way that avoids both the sequent calculus (of which the formulation above is a simplification, using finite sets of formulas instead of sequents) and the use of ordinals. He considers deductions of single formulas in an ordinary Hilbert-style formalization of PA. Assume that with each formula ϕ in the deduction, we associate a certain prenex normal form

$$\exists x_1 \forall y_1 \cdots \exists x_n \forall y_n A(x_i, y_j)$$

of ϕ, with A quantifier-free and denoting a decidable formula. (For example, in $\neg\phi \vee \psi$, bring out all the quantifiers in ϕ first.) Then we may regard sentences of the form $A(k_i, m_j)$ as atomic in playing the game $\mathcal{G}(\{\phi\})$. In this game, if we restrict $\bigwedge$'s moves to those given by functions $f_j(x_1, \ldots, x_j) = y_j$ of the preceding existentially quantified variables, the winning strategy for $\bigvee$ yields functionals $F_1, \ldots, F_n$ such that, taking f to be a code for $(f_1, \ldots, f_n)$,

$$A[F_i(f), f_j(F_1(f), \ldots, F_j(f))]$$

($\bigvee$ gets to let his choices $x_i = F_i(f)$ depend on *all* of the f_j because, as we noted, he can go back at any stage of the game and introduce a new com-

[40] Or, almost. A cut-free deduction which contains an application of a $\bigwedge$-introduction (i.e., a $\wedge$- or a $\forall$-introduction) followed by a $\bigvee$-introduction applied to a different formula does not correspond to a winning strategy: it violates the condition that, when Δ contains a $\bigwedge$-sentence, then $\bigwedge$ moves. But these applications can be commuted, transforming Schütte's cut-free proof into a winning strategy.

ponent of a $\bigvee$-sentence, depending on what $\bigwedge$ has already chosen.) Parsons and Sieg observe in their introductory note that this is the so-called *no-counterexample interpretation*, NCI, first published by Kreisel [1951], who obtained it by means of Ackermann's consistency proof for arithmetic using the ϵ-substitution method [1940]. It is easy to see that, if there is a NCI for a formula, then there is one in *standard form*, i.e. such that

$$\bigwedge_{i<k} f_i(F_1(f), \dots F_i(f)) = g_i(F_1(g), \dots, F_i(g)) \wedge \bigwedge_{i<k} F_i(f) = F_i(g) \longrightarrow \ F_k(f) = F_k(g).$$

To obtain the no-counterexample interpretation directly, we need to obtain the NCI, $F_1, \dots, F_n$, of the conclusion of each inference in the given deduction from the NCI's of its premises. Gödel observes that the crucial case is modus ponens, from ϕ and $\neg\phi \vee \psi$, to ψ. (The case of mathematical induction is obtained from the case of modus ponens by iteration.) To see how he handles this inference, consider the case that ϕ and ψ have the prenex forms $\exists x \forall y \exists z \forall u A(x, y, z, u)$ and $\exists v \forall w B(v, w)$, respectively, so that $\neg\phi \vee \psi$ has the form $\forall x \exists y \forall z \exists u \exists v \forall w[\neg A(x, y, z, u) \vee B(v, w)]$. Thus the NCI's of ϕ and $\neg\phi \vee \psi$ are of the form

$$A[F_1(f), f_1(F_1(f)), F_2(f), f_2(F_1(f), F_2(f)]$$

$$\neg A[g_1, G_1(g), g_2(G_1(g)), G_2(g)] \vee B[G_3(g), g_3(G_1(g), G_2(g)), G_3(g)]$$

where we may assume these are in standard form. We may take for g_3 a function that depends only on its last argument, so that the second formula can be written as

$$\neg A[g_1, G_1(g), g_2(G_1(g)), G_2(g)] \vee B[G_3(g), g_3(G_3(g))]$$

We can solve the equations

$$g_1 = F_1(f),\ f_1(g_1) = G_1(g),\ g_2(f_1(F_1(f))) = F_2(f),\ f_2(g_1, g_2(G_1(g))) = G_2(g)$$

one-by-one from the left, substituting the solution in the subsequent equations, obtaining g_1 as a function of f_1, g_2, f_2, g_3, then f_1 as a function of g_2, f_2, g_3, then g_2 as a function of f_2, g_3 and, finally, f_2 as a function $G(g_3)$ of g_3. Then $B[G(g_3), g_3(G(g_3))]$, a NCI for ψ, is obtained by *modus ponens.* The first equation is an explicit definition. The remaining ones, ignoring parameters and coding pairs of numbers by numbers, have the form

$$h(L(h)) = K(h)$$

where, because the NCI's are in standard form, $L(h) = L(h') \longrightarrow K(h) = K(h')$. We solve these equations one-by-one as follows: define the sequence $\langle h_n \mid n < \omega\rangle$ of approximations to a solution for this equation by $h_0(m) = 0$ for all m and

$$h_{n+1}(m) = \begin{cases} K(h_n) & \text{if } m = L(h_n) \\ h_n(m) & \text{otherwise} \end{cases}$$

It is easy to show that $h_n(m) \neq 0$ implies that $h_n(m) = h_{n+k}(m)$ for all k. We need to infer from this property that, for some n,

$$L(h_n) = L(h_{n+1})$$

For then $h = h_{n+1}$ is the desired solution for the equation. Of course, n itself is a functional $n = N(b)$ of the remaining parameters b.

It is for the existence of such a functional N that Gödel appeals to "Souslin's schema" (a.k.a. Brouwer's Bar Theorem). To describe this, it suffices to consider functionals, such as we have represented L, whose arguments are numerical functions of one numerical value. For, as we already noted, numerical functions of several numerical variables can be represented by one of a single numerical variable, and a functional L of arguments $f_1, \dots, f_m, x_1, \dots, x_n$ can be regarded as a functional of one argument f, where $f(k)$ is the code for the sequence $f_0(k), \dots, f_m(k), x_1, \dots, x_n$. When f is a numerical function, then $\bar{f}(n)$ is the code for the sequence $\langle f(0), \dots, f(n-1)\rangle$. Now the functionals L that we are considering—"functionals that are defined in a finitary way", as Gödel expresses it (p. 111)—have this property: we can associate with them numerical functions Φ_L and Ψ_L and a well-ordering $\prec_L$ of the natural numbers such that

$$\Phi_L(\bar{f}(n)) = 0 \longrightarrow \Psi_L(\bar{f}(n+1)) \prec_L \Psi_L(\bar{f}(n))$$

$$\Phi_L(\bar{f}(n)) \neq 0 \longrightarrow L(f) = \Phi_L(\bar{f}(n+m)) - 1, \text{for all } m.$$

The definition of N is by so-called *Bar Recursion*, which amounts in this case to a nested recursion on $\prec_L$ or a (simple) recursion on $\omega^{\prec_L}$. It is this exponentiation that is responsible for the bound ϵ_0 of the ordinals of the $\prec_H$ for functionals H which occur in the NCI's of formulas in formal deductions in PA.

Finally, to see the connection between the NCI and winning strategies in Gentzen's game, consider the simple case $\exists v \forall w B(v, w)$ and its NCI,

$B[G(f), f(G(f))]$, which we have just considered. Here is the strategy that G determines for $\bigvee$: the latter simply chooses successively

$$\forall w B(0, w), \forall w B(1, w), \ldots, \forall w B(n, w)$$

so that at stage $n + 1$ a set of the form

$$\{\exists v \forall w B(v, w), B(0, b_0), B(1, b_1), \ldots, B(n, b_n)\}$$

is on the board. $\bigvee$ continues with these choices until $\Phi_G(\langle b_0, b_1, \ldots, b_n \rangle) > 0$. At that point it chooses $\forall w B(G(f), w)$, where f is any function with $\bar{f}(n + 1) = \langle b_0, b_1, \ldots, b_n \rangle$. No matter how $\bigwedge$ chooses b_{n+1}, $B(G(f), b_{n+1})$ will be true. Thus, the ordinal of $\prec_G$ is the length of this winning strategy.

References

Ackermann, W. [1940]. Zur Widerspruchsfreiheit der Zahlentheorie, *Mathematische Annalen* **117**: 162–194.

Avigad, J. and Feferman, S. [1998]. Gödel's functional ("Dialectica") interpretation, *[**?**]* pp. 337–405.

Boolos, G. [1971]. The iterative conception of set, *The Journal of Philosophy* **68**: 215–232. In citeboolos:98.

Cantor, G. [1883]. Über unendliche, lineare Punktmannigfaltigkeiten, 5, *Mathematische Annalen* **21**: 545–586. In [**?**].

Cantor, G. [1885]. Rezension der Schrift von G. Frege *Die Grundlagen der Arithmetik*, *Deutsche Literaturzeitung* **6**: 728–729. In [**?**].

Davis, M. [1998]. Review of [**?**]., *Philosophia Mathematica* **6**: 116–128.

Feferman, S. [1997]. Does mathematics need new axioms?, *American Mathematical Monthly* **106**: 99–111.

Frege, G. [1884]. *Grundlagen der Arithmetik*, Breslau: Verlag von Wilhelm Koebner. German/English edition entitled "The Foundations of Arithmetic", Translation into English by J.L. Austin, Oxford: Basil Blackwell (1950).

Friedman, H. [1998]. Finite functions and the necessary use of large cardinals, *Annals of Mathematics* **148**: 803–893.

Gentzen, G. [1936]. Die Widerspruchfreiheit der reinen Zahlentheorie, *Mathematische Annalen* **112**: 493–565.

Gödel, K. [*193?]. Undecidable diophantine propositions, *[Gödel, 1995]*.

Gödel, K. [1933e]. Zur intuitionistischen Arithmetik und Zahlentheorie, *Ergebnisse eines mathematischen Kolloquiums* **4**: 34–38. Reprinted with an Englsh translation in [Gödel, 1986].

Gödel, K. [1933f]. Eine Interpretation des intuitionistischen Aussagenkalküls, *Ergebnisse eines mathematischen Kolloquiums* **4**: 39–40. Reprinted with an Englsh translation in [Gödel, 1986].

Gödel, K. [*1933o]. The present situation in the foundations of mathematics, *[Gödel, 1995]*.

Gödel, K. [*1938a]. Lecture at Zilsel's, *[Gödel, 1995]*.

Gödel, K. [*1939b]. Lecture at Göttingen, *[Gödel, 1995]*.

Gödel, K. [*1940a]. Brown university lecture, *[Gödel, 1995]*.

Gödel, K. [*1941]. In what sense is intuitionistic logic constructive?, *[Gödel, 1995]*.

Gödel, K. [1944]. Russell's mathematical logic, *[?]*, pp. 123–53. Reprinted in [Gödel, 1990].

Gödel, K. [1947]. What is Cantor's continuum problem?, *American Mathematical Monthly* **54**: 515–525. Reprinted in [Gödel, 1990] [Gödel, 1964] is a revised and expanded version.

Gödel, K. [*1951]. Some basic theorems on the foundations of mathematics and their implications, *[Gödel, 1995]*.

Gödel, K. [*1953/9]. Is mathematics syntax of language?, *[Gödel, 1995]*.

Gödel, K. [1958]. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, *Dialectica* **12**: 280–287. Reprinted with an Englsh translation in [Gödel, 1990]. [Gödel, 1972] is a revised version.

Gödel, K. [*1961/?]. The modern development of the foundations of mathematics in the light of philosophy, *[Gödel, 1995]*.

Gödel, K. [1964]. What is Cantor's continuum problem?, *[?]*, pp. 258–273. Revised and expanded version of [Gödel, 1947]. Reprinted in [Gödel, 1990]. In the second edition of [**?**], the pages are 470–485.

Gödel, K. [1972]. On an extension of finitary mathematics which has not yet been used, *[Gödel, 1990]*. Revised version of [Gödel, 1958].

Gödel, K. [1986]. *Collected Works, Vol. I*, Oxford: Oxford University Press.

Gödel, K. [1990]. *Collected Works, Vol. II*, Oxford: Oxford University Press.

Gödel, K. [1995]. *Collected Works, Vol. III*, Oxford: Oxford University Press.

Goodman, N. [1970]. A theory of constructions equivalent to arithmetic, *[?]*, pp. 100–120.

Hanf, W. and Scott, D. [1961]. Classifying inaccessible cardinals (abstract), *Notices of the American mathematical Society* **8**: 445.

Heyting, A. [1930b]. Sur la logique intuitionniste, *Académie Royale de Belgique* **Bulletin 16**: 957–963.

Hilbert, D. [1926]. Über das Unendliche, *Mathematische Annalen* **95**: 161–90. Translated by Stefan Bauer-Mengelberg in [**?**, 367-92].

Hilbert, D. and Bernays, P. [1934]. *Grundlagen der Mathematik I*, Berlin: Springer-Verlag. The second edition was published in 1968.

Howard, W. [1980]. The formula-as-types notion of construction, pp. 479–490.

Kleene, S. [1959]. Recursive functionals and quantifiers of finite types I, *Trans. Amer. Math. Soc.* **91**: 1–52.

Kreisel, G. [1951]. On the interpretation of non-finitist proofs—part i, *The Journal of Symbolic Logic* **16**: 241–267.

Kreisel, G. [1962]. Foundations of intuitionistic logic, *[?]*, pp. 198–210.

Kreisel, G. [1965]. Mathematical logic, *[?]*.

Lévy, A. [1960]. Axiom schemata of strong infinity in axiomatic set theory, *Pacific Journal of Mathematics* **10**: 223–238.

Parsons, C. [1995a]. A distinction in platonism, *Under Tall Trees: A Tribute to Dirk van de Kaa* pp. 9–10.

Parsons, C. [1995b]. Platonism and mathematical intuition in kurt gödel's thought, *Bulletin of Symbolic Logic* **1**: 44–74.

Parsons, C. [n.d.]. realism and the debate on impredicativity, 1917–1944. To appear in the proceedings of the conference *Reflections* held on December 13–15, 1998, at Stanford University.

Reinhardt, W. [1974]. Remarks on reflection principles, large cardinals, and elementary embeddings, *Proceedings of Symposia in Pure Mathematics 10. Part II*, pp. 189–205.

Schönfinkel, M. [1924]. Über die Bausteine der mathematischen Logik, *Mathematische Annalen* **92**: 305–316.

Schütte, K. [1951]. Beweisetheoretische Erfassung der unendlichen Induktion in der Zahlentheorie, *Mathematische Annalen* **122**: 369–389.

Schütte, K. [1965]. Eine Grenze für die Beweisebarkeit der transfinitien Induction in der verzweigten Typenlogik, *Archiv für mathematische Logik und Grundlagenforschung* (7).

Scott, D. [1961]. Measurable cardinals and constructible sets, *Bulletin de l'Académie Polonaise des Sciences, Série des Sciences Mathematiues* **9**: 521–524.

Shoenfield, J. [1967]. *Mathematical Logic*, London: Addison-Wesley.

Spector, C. [1962]. Provably recursive functionals of analysis: a consistency proof of analysis by an extension of the principles formulated in current intuitionistc mathematics, *[?]*.

Tait, W. [1963]. A second order theory of functionals of higher type, with two appendices. Appendix A: Intensional functionals. Appendix B. An interpretation of functionals by convertible terms., *Stanford Seminar Report* pp. 171–206. A published version is [Tait, 1967].

Tait, W. [1967]. Intensional interpretations of functionals of finite type i, *Journal of Symbolic Logic* **32**: 198–212.

Tait, W. [1981]. Finitism, *Journal of Philosophy* **78**: 524–556.

Tait, W. [1986]. Truth and proof: the Platonism of mathematics, *Synthese* **69**: 341–370.

Tait, W. [2001]. Beyond the axioms: the question of objectivity in mathematics, *Philosophia Mathematica* **9**: 21–36.

Tait, W. [n.d.]. Remarks on finitism. To appear in the proceedings of the conference *Reflections* held on December 13–15, 1998, at Stanford University.

von Neumann, J. [1929]. Über eine Widrspruchsfreiheitsfrage in der axiomatischen Mengenlehre, *Journal für die reine und angewandte Mathematik* **160**: 227–41.

Wang, H. [1974]. *From Mathematics to Philosophy*, London:Routledge.

Wang, H. [1987]. *Reflections on Kurt Gödel*, Cambridge: MIT Press.

Wang, H. [1996]. *A Logical Journey: From Gödel to Philosophy*, Cambridge: MIT Press.

Zach, R. [1998]. Numbers and functions in Hilbert's finitism, *Taiwanese Journal for Philosophy and History of Science* **10**: 33–60.

Zermelo, E. [1930]. Über Grenzzahlen und Mengenbereiche. Neue Untersuchungen über die Grundlagen der Mengenlehre, *Fundamenta Mathematicae* **16**: 29–47. Translated in [**?**, 1219-1233].